Doctoral Thesis

Nonparametric inference for first-order characteristics of spatial and spatio-temporal point processes

Isabel Fuentes Santos

Departamento de Estatística e Investigación Operativa

Universidade de Santiago de Compostela

November 2015

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Nonparametric inference for first-order characteristics of spatial and spatio-temporal point processes

Isabel Fuentes Santos

Supervisors
Dr. Wenceslao González Manteiga
Dr. Jorge Mateu Mahiques

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Informe de los Directores

Don Wenceslao González Manteiga, Catedrático del Departamento de Estatística e Investigación Operativa de la Universidade de Santiago de Compostela y Don Jorge Mateu Mahiques, Catedrático del Departamento de Matemáticas de la Universidad Jaume I de Castellón, informan que la memoria titulada

Nonparametric inference for first-order characteristics of spatial and spatio-temporal point processes

ha sido realizada bajo su dirección por Doña Isabel Fuentes Santos, estimando que la interesada se encuentra en condiciones de optar al grado de Doctor, por lo que se solicita que sea admitida a trámite para su lecture y defensa pública.

Santiago de Compostela, 9 de noviembre de 2015

Los directores:

Prof. Dr. Wenceslao González Manteiga Prof. Dr. Jorge Mateu Mahiques

La doctoranda:

Isabel Fuentes Santos

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

A mis padres. José y Choncha

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

"Statistics, the science of uncertainty,
attempts to model order in disorder"

Noel Cressie

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Agradecimientos

En primer lugar quiero expresar mi más profundo agradecimiento a mis directores, Wenceslao González Manteiga y Jorge Mateu Mahiques, por su asesoramiento científico, su apoyo y, sobre todo su confianza e infinita paciencia cada vez que he entrado en bucle infinito e incumplido plazos. A Wences le agradezco que me haya aceptado como estudiante de doctorado, que me haya ofrecido trabajar con procesos puntuales, un verdadero regalo, y que me haya transmitido su entusiasmo y motivación. A Jorge le agradezco su implicación desde el primer momento, desde la invitación al SPPA que me brindó la oportunidad de descubrir los procesos puntuales de la mano de sus máximos exponentes, su hospitalidad en las visitas a Castellón, y su apoyo a través del mail.

Gracias a Rosa Crujeiras y Jesús Aboal por haberme cedido los datos de Laurisilva para el proyecto del DEA y por haber seguido contando conmigo. A Rosa también quiero agradecerle su asesoramiento metodológico en la primera etapa de este largo camino. Gracias a Manuel Marey por compartir conmigo los datos de incendios forestales y todo sus conocimientos sobre ese gran problema. Gracias a Maribel por su colaboración en esta última etapa. Agradezco a Manuel febrero su ayuda con nuestro amigo R. Gracias a todos los miembros del departamento de Estadística e Investigación Operativa por haberme echado una mano cada vez que lo he necesitado.

Quiero agradecer a los miembros del área de Estadística e Investigación Operativa de la UDC su acogida, apoyo y todos los buenos momentos compartidos durante mis dos años de docencia en ese departamento. Gracias a Ricardo Cao por su asesoramiento con el bootstrap suavizado. Álex, acerté al 100% al elegir compañero de fatigas docentes, Yolanda, echo de menos nuestras charlas de sobremesa y tren. Asimismo, agradezco el apoyo recibido por los miembros del departamento de Estadística e Investigación Operativa en el campus de Ourense durante mi etapa como profesora visitante.

A mis jefes del grupo ESMABA en el Instituto de investigaciones Marinas, Mº José Fernández Reiriz y Uxío Labarta les agradezco la oportunidad de trabajar en un grupo multidisciplinar y las facilidades que me han dado para compatibilizar mi trabajo con la elaboración de esta tesis. En particular, quiero agradecer a Uxío su apoyo e insistencia para que terminase la tesis. Muchas gracias también a mis compañeros Bea, Elsi, Jose, Kris y Lourdes, por los ánimos, las charlas, los cafés ..., Eva, ánimo que no te queda nada.

No hay palabras para expresar mi agradecimiento a los principales responsables de que yo haya llegado hasta aquí: José y Choncha, mis padres. Sin vuestro esfuerzo, sacrificio, confianza y apoyo esto no habría sido posible. Gracias papá, por fomentar mi curiosidad desde pequeñita e intentar encontrar respuesta a todas mis preguntas. Gracias mamá, por los consejos, por ser un ejemplo de fuerza y superación y enseñarme que hay que seguir luchando aunque todo indique que la guerra está perdida. Quiero agradecer a mi familia su apoyo incondicional, muchas gracias por haberos multiplicado para ayudarnos durante los malos momentos permitiéndome seguir con este proyecto. En especial gracias a mi madrina, mi tía Avelina y mi tío Alejandro por hacerme sentir que parte de ellos sigue conmigo.

No puedo cerrar estos agradecimientos sin recordar a mis amigos, que me han acompañado, animado y, en "contadas" ocasiones, sufrido durante este largo viaje. Isa y Rocío, que están conmigo desde la carrera; Pamela, que llegó a mi vida a la vez que los procesos puntuales y ha llegado a ser incluso más importante que ellos; Alhambra, que ha tenido que volverse de la Postdoc sin que haya podido ir a visitarla; Fandi, que sufre un sarpullidos cuando oye la palabra tesis; Primo, que puede hacerme sonreír incluso en los momentos más difíciles; Álex y sus buenos consejos; Omar, que me ha salvado de alguna crisis informática; Kike, que me ha ayudado a desconectar del estrés; y los compañeros de baile. Gracias a todos por el apoyo en los malos momentos y por hacer que los buenos sean aún mejores.

Finalmente hago constar que este trabajo ha sido financiado por los proyectos "MTM2008-0310: Metodologías y aplicaciones en estadística semiparamétrica, funcional y espacio-temporal" del Ministerio de Ciencia e Innovación, "MTM2013-41383-P: Inferencia no paramétrica: modelización, estimación, contrastes y aplicaciones" y "MTM2013-43917-P: Nuevas familias de procesos puntuales espacio-temporales con intensidades procedentes de campos aleatorios" del Ministerio de Economía y Competitividad, y "P1-1B2012-52: Características locales de segundo orden en procesos estocásticos espacio-temporales. Aplicaciones en Epidemiología y Medio Ambiente" de la Fundación Bancaja.

Santiago de Compostela, 5 de noviembre de 2015

Isabel Fuentes Santos

Contents

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Preliminaries

This dissertation focuses on the analysis of spatial and spatio-temporal point processes. Point processes are mathematical models that describe the arrangement of objects randomly distributed in the plane or the space, Any realization of a point process is referred as a point pattern. To find the roots of point process we should go back to Poisson in 1837. However, much of the point process theory has been developed during the last decades, in parallel with a profound theoretical development in other areas of statistics, and enhanced by the increasing number of scientific fields dealing with data irregularly distributed in the space or in the spatio-temporal domain (e.g. archeology, astronomy, forestry, biology, ecology, epidemiology, seismology). Theoretical background for the analysis of point processes can be found in Daley and Vere-Jones (2003), Moller and Waagepetersen (2003), Daley and Vere-Jones (2007), Illian et al. (2008), and Diggle (2013).

One of the main issues in the analysis of spatial and spatio-temporal point processes is to model their first-order intensity function, i.e. to describe their distribution in the observation domain. Intensity estimation can be addressed from a parametric or from a nonparametric perspective. The parametric approach assumes a model for the intensity function and the problem reduces to estimate its unknown parameters. However, we can obtain unreliable estimates when the assumed model deviates from the true intensity. The nonparametric approach relies on the observed data without assuming any restrictive model and "let the data speak for themselves". Following the ideas of kernel density estimation, Diggle (1985) introduced the nonparametric kernel estimator of the first-order intensity function for point processes in $\mathbb{R}$, which can be directly extended

to the spatial and spatio-temporal domains, However, the lack of consistency of this estimator has limited its use to the exploratory analysis, while in other areas of statistic nonparametric techniques have provided an increasing body of powerful inference tools. In order to overcome this drawback Cucala (2006) introduced the density of event locations and proved the consistency of its kernel estimator.

The selection of the bandwidth parameter is a critical issue in any nonparametric procedure that has received little attention in the point process context. For stationary Cox processes in $\mathbb{R}$ Diggle and Marron (1988) found that the bandwidth minimizing the mean squared error (MSE) of the kernel intensity estimator equals that provided by least-squares cross-validation (LSCV) for the kernel density estimator. Brooks and Marron (1991) showed that the LSCV-optimal bandwidth is asymptotically optimal for inhomogeneous point processes in $\mathbb{R}$. Other criteria such as plug-in and bootstrap bandwidth selectors, which good performance has been proved in other areas of statistics, has not been addressed in kernel intensity estimation. In should also be noted that, although spatial point processes arise in the plane, scalar bandwidth has been commonly used to estimate the first-order intensity function.

Modeling the joint distribution of spatial locations and times of occurrence in spatio-temporal point processes can be a difficult task, whose complexity increases when the point process is marked or depends on covariates. For this reason, most of the current models assume that the spatio-temporal intensity function is separable, i.e. it can be expressed as the product of its spatial and temporal components. However, this assumption can be very restrictive and unrealistic in practice. This situation has motivated the development of nonparametric separability tests based on Monte Carlo simulations of the separable model (Schoenberg, 2004; Assuncáo and Maia, 2007; Chang and Schoenberg, 2011; Díaz-Avalos et al., 2013).

1.2 Real data problem

1.2.1 Motivation

Wildfire is the most ubiquitous natural disturbance in the world and represents a problem of considerable social and environmental importance. A wildfire is any uncontrolled fire in combustible vegetation in the countryside or in a forestry area. A wildfire differs from other fires in its extensive size, the speed at which it can spread out from its

original source, its potential to change direction unexpectedly, and its ability to jump gaps such as roads, rivers and fire breaks. Wildfires can be characterized according to the cause of ignition, physical properties such as speed of propagation, the type of combustible material, or the effect of weather on the fire. The four major natural causes of wildfire ignition are lightning, volcanic eruption, sparks from rock falls, and spontaneous combustion (National Interagency Fire Center., 2011). However, many wildfires are attributed to human sources (Pyne et al., 1996).

In Galicia arson fires, which during the last decades have represented more than the 75% of wildfires registered in this region, are the main cause of forest destruction and, consequently, one of main concerns faced by the Government and environmental managers. This situation has motivated an increasing interest in the development of statistical techniques to understand the behavior of wildfire, which would be of great value to develop effective prevention and fire fighting plans.

If we associate wildfires to their spatial coordinates, the latitude and longitude of the ignition point, and time of detection, along with other variables such as cause, size, or type of burned area, they can be seen as a realization of a spatio-temporal point process. The analysis and inference techniques recently developed for spatial and spatio-temporal point processes can provide valuable information regarding the spatial distribution of wildfires, the dependence between them and their dependence on environmental and weather conditions. During the last decade several works have used spatial and spatio-temporal point process methods to analyze wildfires risk (Fuentes-Santos et al., 2013; Genton et al., 2006; Hering et al., 2009; Juan et al., 2012; Møller and Díaz-Avalos, 2010; Pereira et al., 2013; Schoenberg et al., 2007; Serra et al., 2014; Turner, 2009; Xu et al., 2011; Yang et al., 2007).

We shall apply analysis and inference techniques currently available for point processes, as well as those introduced in this thesis, to the spatio-temporal pattern of wildfires registered in Galicia during the decade $1999 - 2008$, which comprises a total amount of 85134 ignition points classified by burned area and cause of fire. In Section 1.2.2 we provide a detailed description of the wildfire dataset.

1.2.2 Wildfires dataset

The dataset comprises the 85134 wildfires recorded in Galicia in the period 1999-2008, which affected 318861 ha of forested land, equivalent to the 15% of the total forest area in Galicia, and 5014 ha of non-forested land. Taking into account the classification done in the PLADIGA 2015 (`http://www.medioruralemar.xunta.es/areas/forestal/incendios_forestais/pladiga_2015`), we define three types of wildfires according to the burned area (S): small fires ($S < 1$ ha), regular fires ($1 \leq S < 25$ ha), and large fires ($S \geq 25$ ha). Figure 1.1 shows that the majority of wildfires recorded in Galicia during the study period burned less than 1 ha, small fires represented between the 69.5% of wildfires in 2006 to the 77.2% in 1999. While the percentage of large wildfires was between the 0.6% in 1999 and the 4.65% in 2006. This implies that a small proportion of wildfires burned most of the affected area. Indeed, the large wildfires registered in August 2006 burned more than the 25% of the total area burned during this decade. The large amount of small fires is one of the particular features of Galicia, as shows the comparison between Galicia and the rest of Spain (Figures 1.2 and 1.3). The number of fires in Galicia can represent more that the 50% of the total of wildfires registered in Spain, but the proportion of affected area used to be much lower.

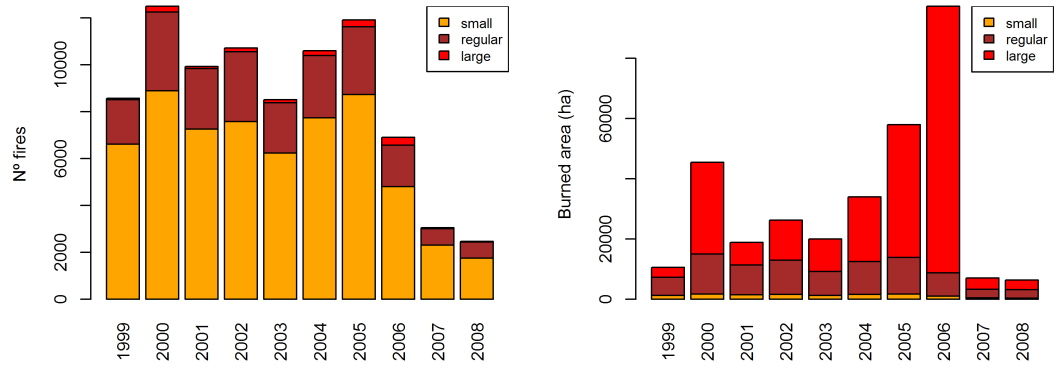


FIGURE 1.1: Number of wildfires (left) and burned area (right) by year and size: small ($S < 1$ ha), regular wildfires ($1 \leq S < 25$ ha) and large ($S \geq 25$ ha).

Wildfires were also classified according to their cause into arson, natural, negligence, reproduction and fires with unknown cause. Figure 1.4 shows the high incidence of human-caused wildfires, including intentionally caused fires and negligences, in Galicia. Arson fires, which represented more than the 80% of wildfires during the study period, are the main cause of forest destruction in Galicia. In the rest of Spain negligence has been the first cause of fire during the last decades, while the percentage of arson fires

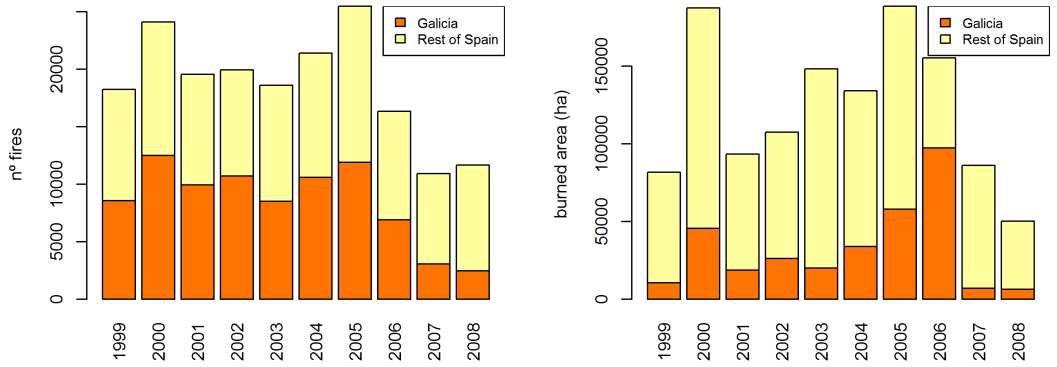


FIGURE 1.2: Number of wildfires (left) and burned area (right), comparison between Galicia and the rest of Spain (1999-2008). From Estadística de Incendios, Ministerio de Agricultura, Alimentación y Medio Ambiente (`http://www.magrama.gob.es/es/biodiversidad/estadisticas/Incendios_default.aspx`)

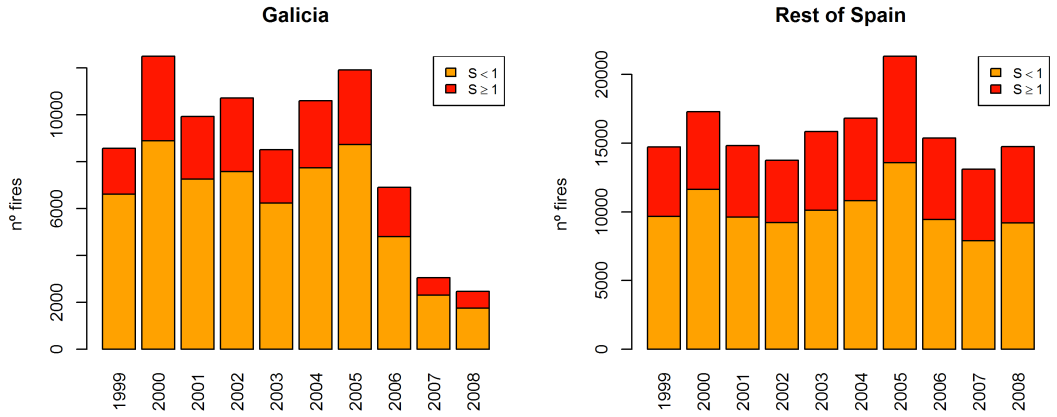


FIGURE 1.3: Proportion of wildfire with $S < 1$ and $S \geq 1$, comparison between Galicia and the rest of Spain (1999-2008).

was bellow 30% of the total.

In the original dataset the spatial locations of ignition points were defined by parish and use of land. In order to treat these data as a spatial point pattern we need to associate spatial coordinates to each ignition point. The alphanumerical information associated to each ignition point was translated to the actual land area with the aid of Geographic Information Systems (GIS) by the procedure described in Figure 1.5 to obtain the corresponding UTM (Universal Transverse Mercator) coordinates. Once

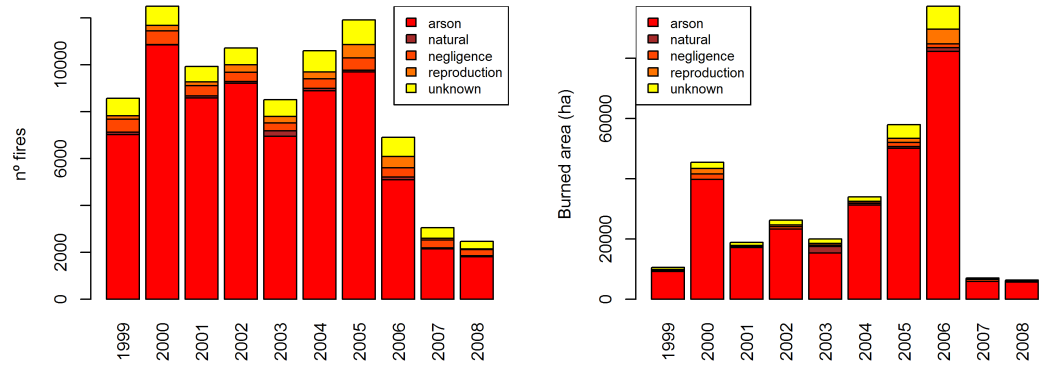


FIGURE 1.4: Number of wildfires (top) and burned area (bottom) by year and cause

identified the spatial coordinates of the ignition points, we have the following spatio-temporal marked point process

$$\mathbf{S} = \{(\mathbf{x_i}, \mathbf{t_i}, size_i, cause_i); i = 1, \ldots, 85134\} \tag{1.2.1}$$

where $\mathbf{x_i}$ are the spatial coordinates of each ignition point, $\mathbf{t_i}$ the date and time of detection, $size_i$ the size group according to the classification outlined above, and $cause_i$ the cause of fire.

Figures 1.6 and 1.7 show the spatial patterns of wildfires and their distribution by size for the ten year under study. A first visual inspection shows the low incidence of large fires, and suggest that the risk of forest fires is higher in the southern area of Galicia than in the Northern-East.

Figures 1.8 and 1.9, which show the spatial patterns of ignition point by cause for each year, highlight the high incidence of arson wildfires, in contrast with the low risk of fires with natural cause. The observed patterns suggest a higher incidence of natural wildfires in Lugo and Ourense than in the West coast of Galicia. Visual inspection does not provide insight into the spatial distribution of human-caused (arson and negligences) wildfires.

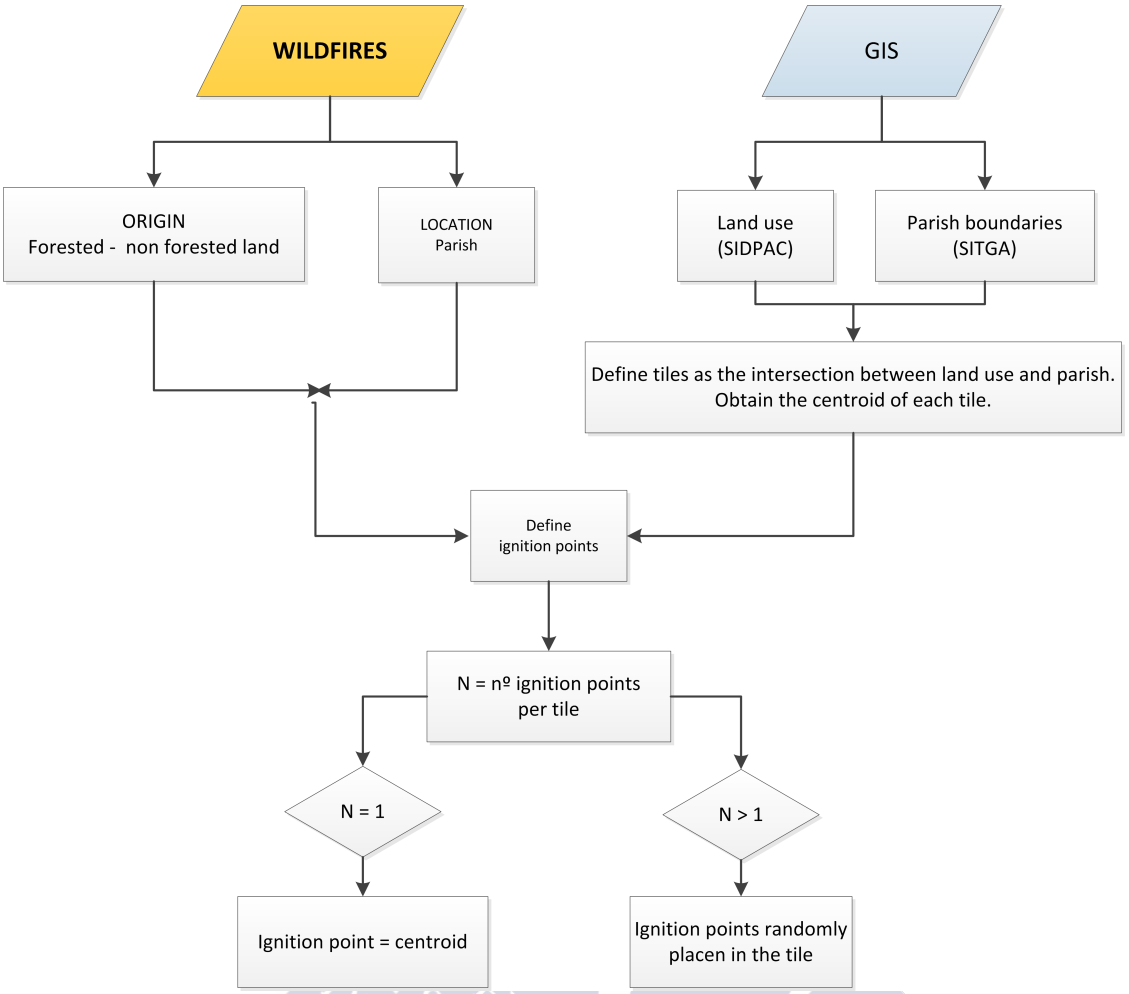


FIGURE 1.5: Diagram of the procedure conducted to define the spatial coordinates of ignition points

1.3 Aims and structure of the thesis

This thesis seeks to provide some methodological advances for the two issues outlined in Section 1.1 and to illustrate their applicability in the analysis of the wildfire dataset introducend in Section 1.2. For this purpose, we have investigated the conditions under which nonparametric inference and smooth bootstrap techniques recently developed for univariate and multivariate data can be extended to the spatial and spatio-temporal point process frameworks. Our work can be divided into four main aims:

1. Develop a smooth bootstrap procedure for inhomogeneous spatial Poisson point processes in order to provide a consistent approach for the mean integrated square error (MISE) of the kernel density of event locations, and propose a procedure based on the MISE bootstrap to select the bandwidth matrix of the kernel intensity estimator.

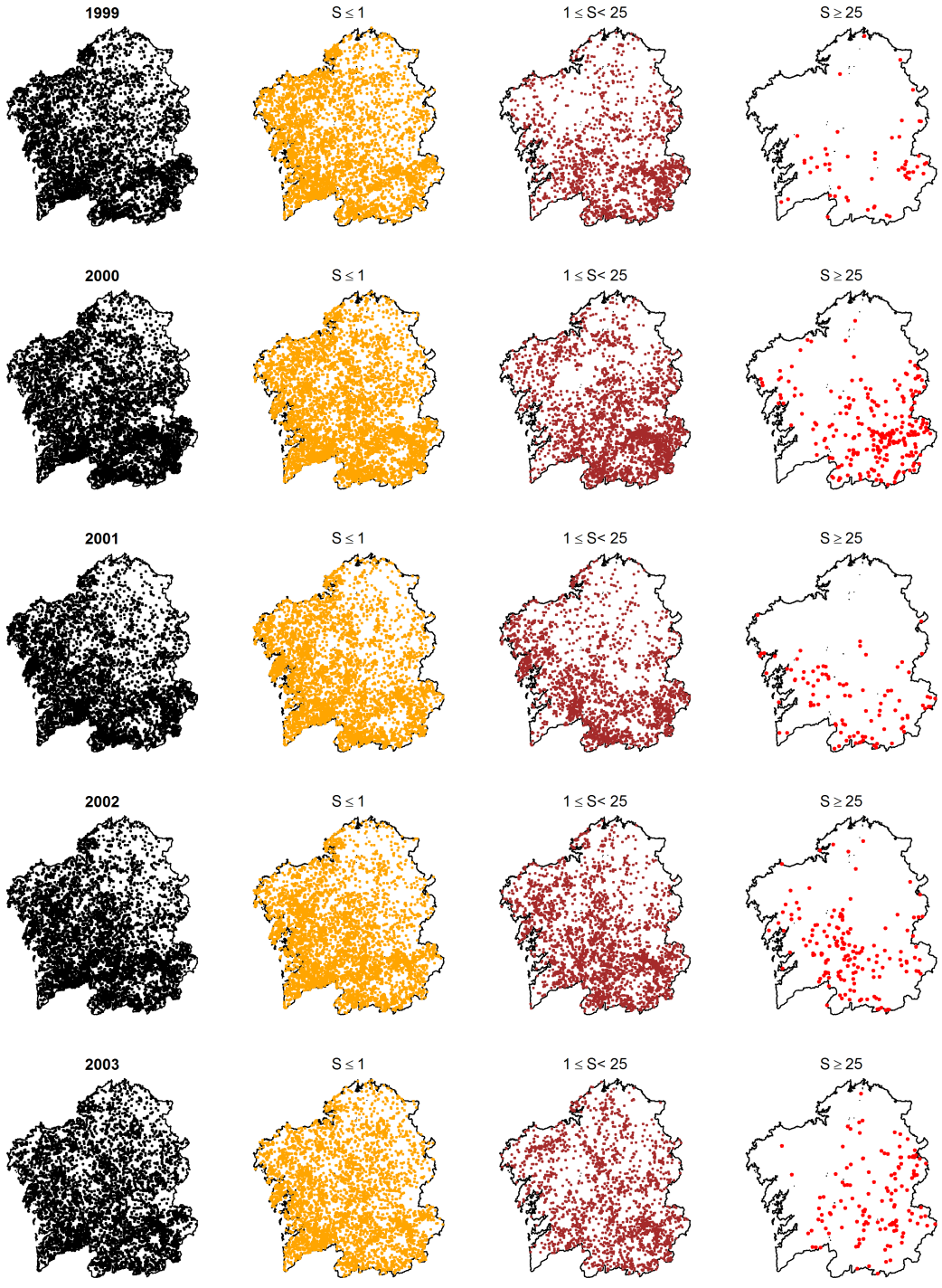


FIGURE 1.6: Spatial pattern of wildfires, small wildfires ($S < 1$ ha), regular wildfires ($1 \leq S < 25$ ha) and large ($S \geq 25$ ha) registered in Galicia from 2004 to 2008

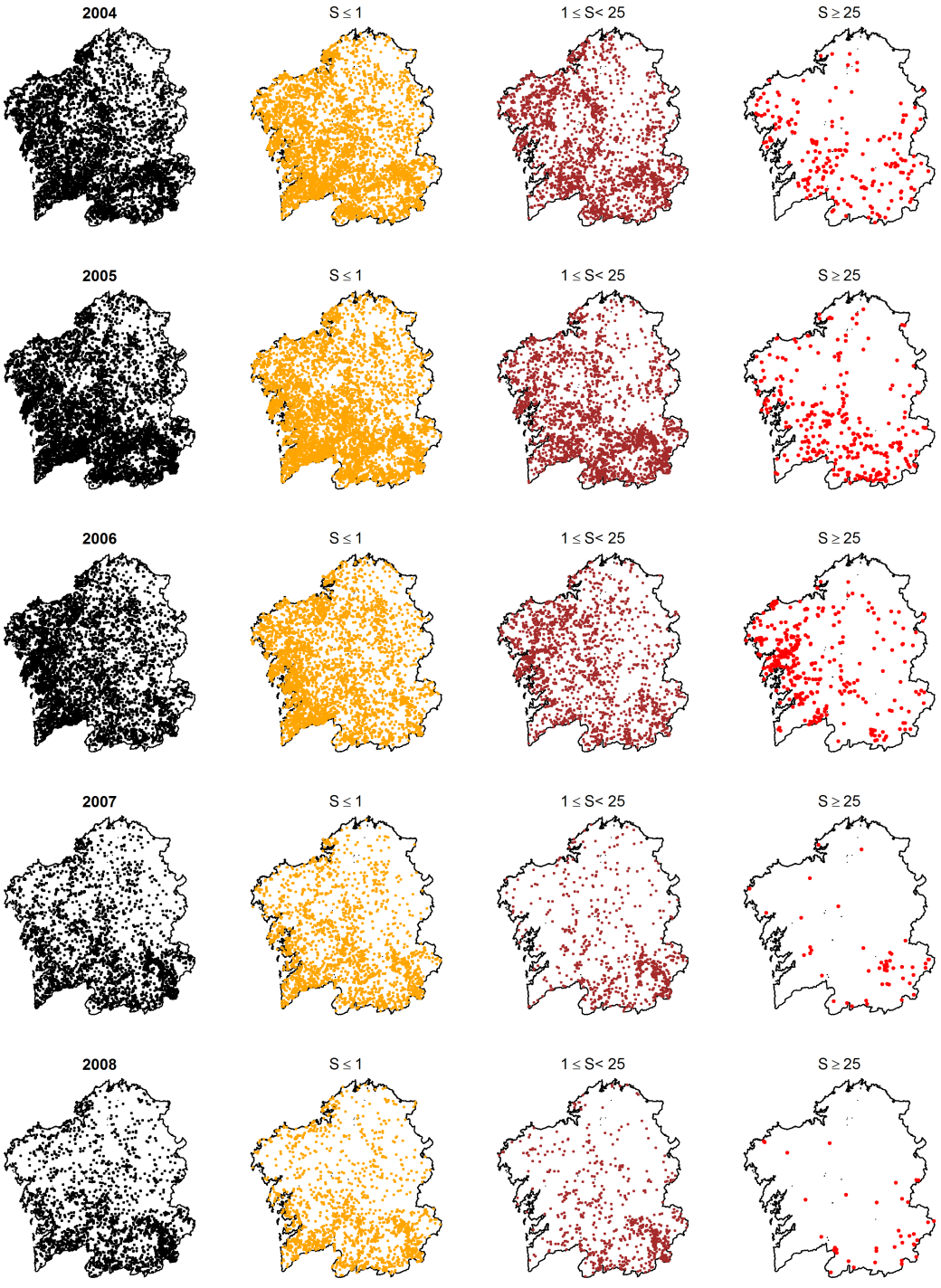


FIGURE 1.7: Spatial pattern of wildfires, small wildfires ($S < 1$ ha), regular wildfires ($1 \leq S < 25$ ha) and large ($S \geq 25$ ha) registered in Galicia from 2004 to 2008

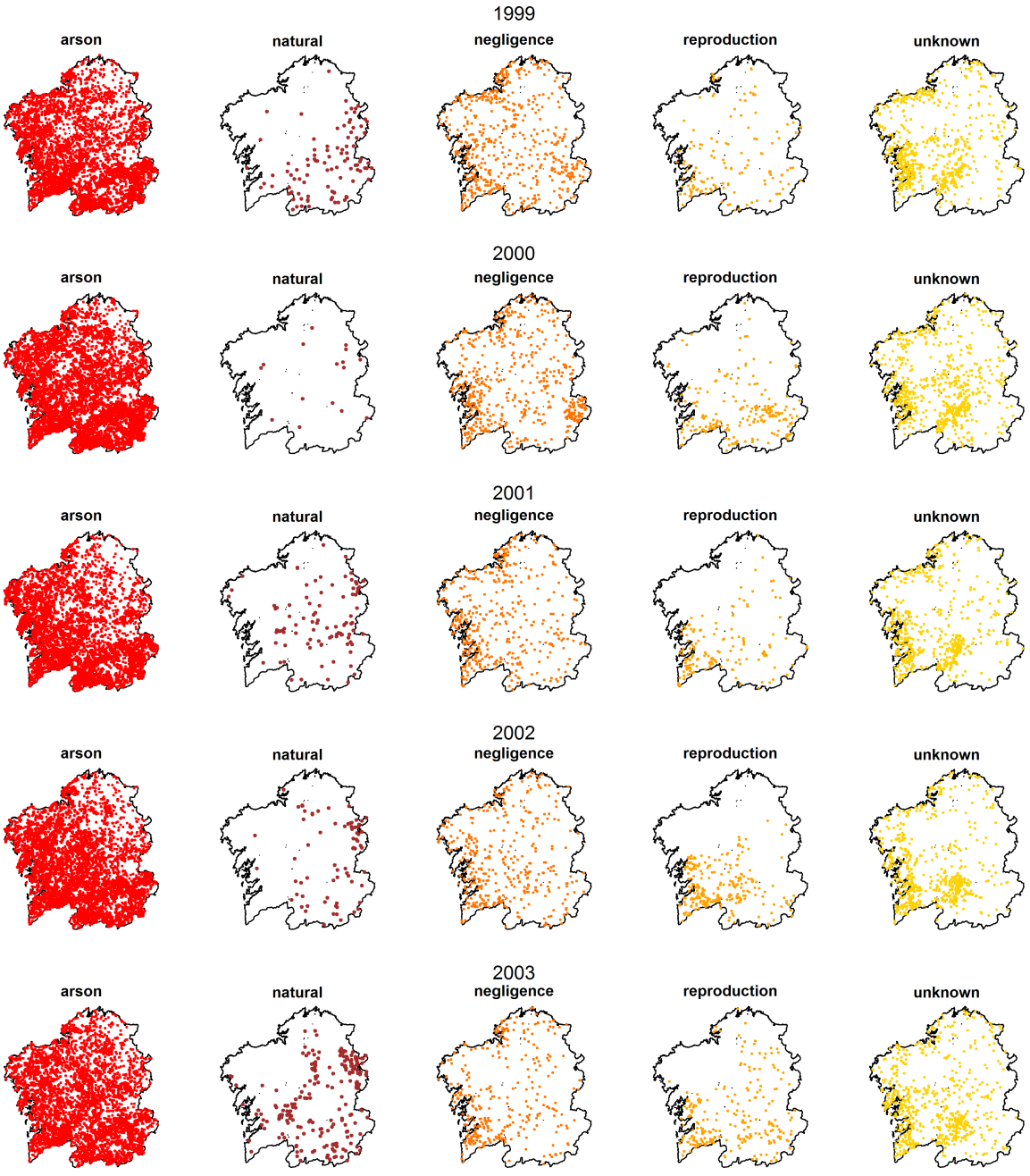


FIGURE 1.8: Spatial pattern of wildfires registered in Galicia from 1999 to 2003 by cause

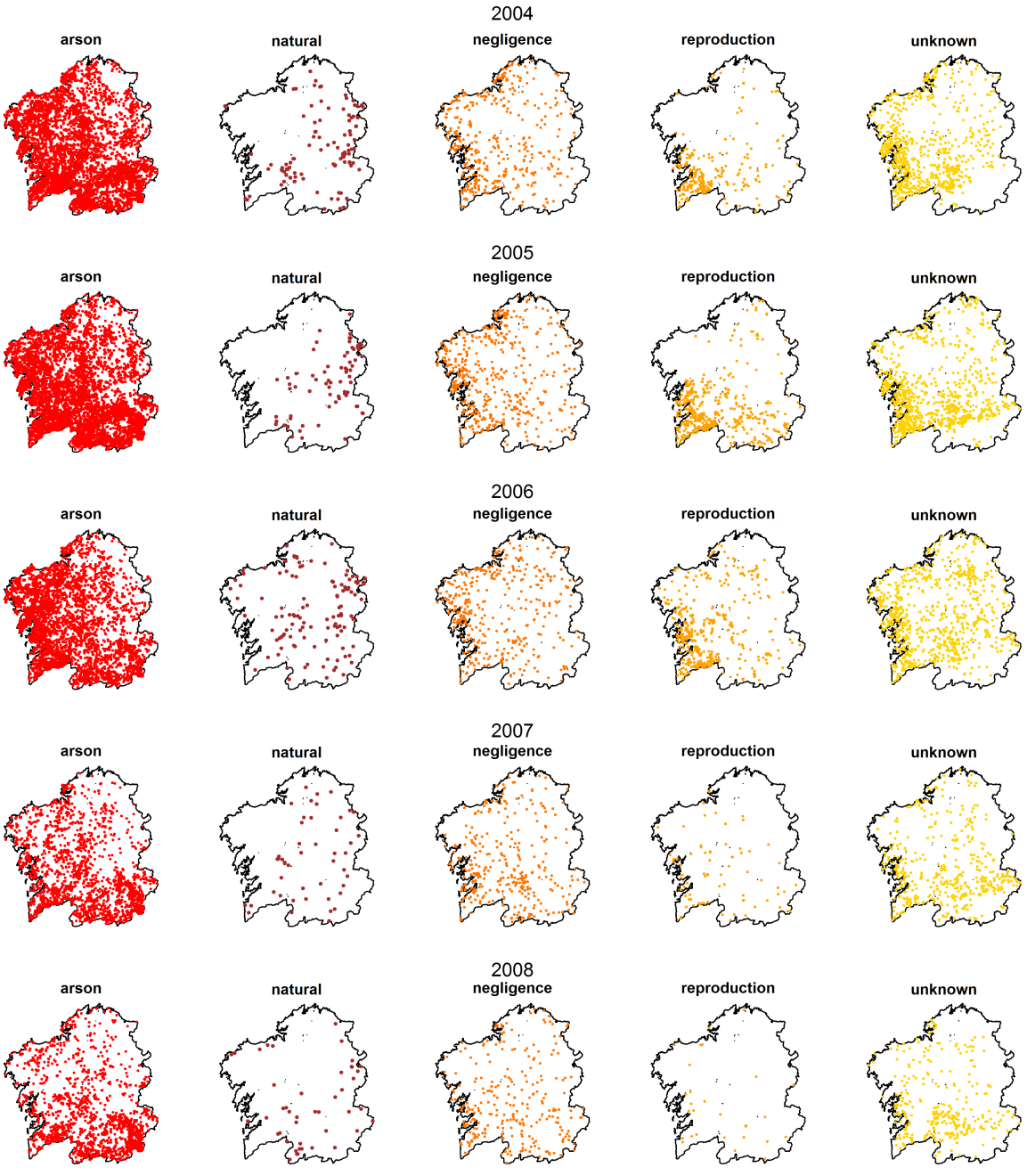


FIGURE 1.9: Spatial pattern of wildfires registered in Galicia from 2004 to 2008 by cause

2. Develop a nonparametric test to compare the intensity functions of inhomogeneous spatial point processes.

3. Develop a nonparametric separability test for spatio-temporal point processes whose calibration does not rely on Monte Carlo simulations of separable point processes.

4. Use the methodology currently available and the techniques introduced in this work to analyze the spatial and spatio-temporal patterns of the wildfires registered in Galicia.

In agreement with these aims, the remainder of this thesis is organized as follows:

- Chapter 2 provides a review of the theory of spatial and spatio-temporal point processes. We introduce the main properties of point processes, as well as their first and second-order characteristics.

- Chapter 3 discusses the kernel intensity estimator and the bandwidth selector procedures currently in use for spatial point processes. We analyze the consistency of the kernel estimator of the density of event locations with bandwidth matrices, in contrast with the scalar bandwidth commonly used. Then we introduce the smooth bootstrap procedure, show the consistence of the bootstrap MISE, and propose a procedure to select the optimal bandwidth matrix. Finally, the performance of the new bandwidth selector is checked on a simulation study and by its application to the wildfires dataset. Supporting information for this chapter is provide din Appendix A.

- Taking into account the findings of Chapter 3, Chapter 4 analyzes whether the nonparametric test proposed by Duong et al. (2012) for comparison of multivariate data, can be extended to compare the first-order structure of spatial point patterns. We propose a bootstrap calibration for the null distribution of the test statistic, and check the performance of the test on simulated data and on the wildfires dataset.

- In Chapter 5 we analyze the behavior of the wildfires registered in Galicia from 1999 to 2008, with the aid of the statistical techniques available for spatial point processes and the new methods developed in Chapters 3 and 4. We analyze the spatial structure of wildfires, as well as the dependence between ignition points.

Supporting information for this chapter is provided in Appendix B, this appendix contain the first-order intensities of the spatial patterns of wildfires by size and cause for each year, and the respective inhomogeneous L-tests.

- Chapter 6 addresses the spatio-temporal separability assumption. We propose using a nonparametric regression tests that checks whether the ratio between the spatio-temporal and spatial first-order intensity functions depend on the spatial locations of events, to test the separability assumption without the need of simulating separable point processes. To implement the test, we first introduce the kernel estimator of the log-ratio function, and propose a bandwidth selector. We analyze the performance of the test in a simulation study and throughout its application to the wildfires dataset. Supplementary information for this chapter is provided in Appendix C.

- Chapter 7 outline some ideas for future research in the issues addressed along the thesis.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 2

A short review on the analysis of point processes

2.1 Introduction

Data irregularly distributed within a region of the one, two or three-dimensional space arise in a wide variety of scientific contexts including ecology, forestry, seismology, epidemiology, cosmology and geography. These datasets are known as *point patterns* and we relate to the locations as *events*, to distinguish them from arbitrary points in the observation domain. The aim of point process statistics is to analyze the structure of these patterns.

A *point process* is a stochastic process that generates a random collection of *events*, $\{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$, in some metric space (Diggle, 2003; Illian et al., 2008). To ease notation, throughout this thesis point processes and patterns are denoted in bold capitals, while the events are denoted in bold. Point processes are characterized by the probability function $P(N(W) = n)\cdot$, where $N(W) = \#(\mathbf{x_i} \in W)$, which is the probability of finding n events in the region W, and by their first and second-order characteristics. The first-order characteristics describe the spatial distribution of events, while the second-order characteristics describe the dependence structure of the point process.

A *spatial point process* generates events on a bounded region $W \subset \mathbb{R}^2$, which defines a bonded support for the characteristic of the point process and introduces *edge-effects*. If the spatial coordinates of the events have attached any measure or *mark*, we have a *marked spatial point pattern*. If the event observed at location $\mathbf{x} \in \mathbb{R}^2$ has associated a

mark, $z \in F$, then $(\mathbf{x}, z)$ is a point in $\mathbb{R}^2 \times F$. Thus a marked spatial point pattern can be seen as realization of a point process defined in $\mathbb{R}^2 \times F$. Finally, if in addition to the spatial location of the events, $\mathbf{x_i} \in W \subset \mathbb{R}^2$, we know the time of occurrence, i.e. the events are observed at times $\mathbf{t_i} \in T \subset \mathbb{R}^+$, we have a *spatio-temporal point pattern*.

This Chapter provides a short review of point process theory. We introduce the main properties and characteristics of spatial and spatio-temporal processes that shall be used throughout this thesis.

2.2 Spatial point processes

A spatial point process is a stochastic process that generates events, $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$, on a bounded region $W \subset \mathbb{R}^2$. In this section we introduce the main characteristics and properties of spatial point processes, which can be directly extended to the multitype and spatio-temporal frameworks.

2.2.1 Complete spatial randomness

Complete spatial randomness (CSR hereafter) is the white noise assumption in point process theory, as it characterizes the lack of structure in the observed pattern. A completely random spatial point process (Figure 2.1, left), which is known as an *homogeneous Poisson process (HPP)*, is characterized by the following properties:

HPP1 the expected number of events in any planar region, W, with surface area $|W|$ has Poisson distribution with mean $\lambda|W|$,

HPP2 given n events in the region W, the spatial locations of these events, $\mathbf{x_i}, i = 1, \ldots, n$, are an independent and random sample of the uniform distribution on W.

The constant λ in condition (HPP1) is the intensity of the process, or expected number of events in W, according to (HPP1) CSR implies that the intensity of the process does not depend on the spatial locations, i.e. the point process is *first-order stationary* or *homogeneous*. According to (HPP2) CSR also implies that there are no interactions among events, i.e. the point process is *Poisson*. The independence assumption is violated when the presence of an event at a given location, x, favors or inhibits the occurrence of other

events around x, given rise to clustered (Figure 2.1, centre) and regular (Figure 2.1, right) point processes, respectively.

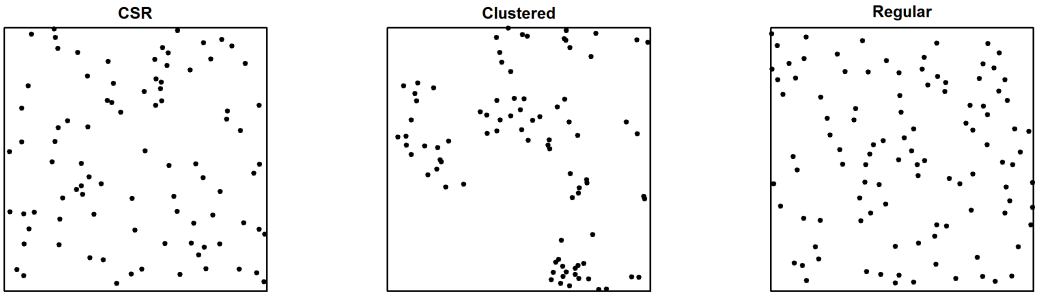


FIGURE 2.1: Realization of completely random, clustered and regular spatial point processes on the unit square with first order intensity $\lambda = 100$.

Testing CSR is the first step in the analysis of any observed spatial point pattern for two main reasons: if the CSR assumption is accepted the point process does not need any further analysis; and CSR acts as a dividing hypothesis that allow us to classify between regular and aggregated patterns. In Section 2.2.6 we introduce the procedures commonly used to test the CSR hypothesis.

2.2.2 Monte Carlo tests

Even simple stochastic models for spatial point patterns lead to intractable distribution theory. This situation motivates the extensive use of *Monte Carlo tests*. Let u_1 the observed value of a statistic U and $u_i; i = 2, \ldots, B$, the corresponding values generated by independent random sampling from the distribution of U under a simple null hypothesis, $\mathcal{H}_0$. Let $u_{(j)}$ be the jth largest value among $u_i; i = 2, \ldots, B$. Thus, under $\mathcal{H}_0$,

$$P\left(u_1 = u_{(j)}\right) = B^{-1},\ j = 2, \ldots, B$$

assuming non-duplicated u_is, rejection of $\mathcal{H}_0$ when u_1 ranks the kth largest or higher provides a one-sided test of size k/B. The extension to two-sided tests is trivial.

Hope (1968) showed through several examples that the loss of power resulting from Monte Carlo implementation is slight, so that a large B is not necessary. For instance, $B = 100$ is sufficient to conduct a one sided test with significance level $\alpha = 0.05$. Power loss in Monte Carlo tests is related to "blurred critical regions" (Diggle, 2003; Marriott, 1979),which arise when the Monte Carlo tests does not find significant values of u_i that are declared significant by classical tests, and vice versa.

An inherent weakness of Monte Carlo tests is their restriction to test simple null hypothesis. Composite hypothesis could be tested generating pseudo-random samples conditional to sufficient statistics, but this is not feasible in practice. For instance, goodness-of-fit tests that ignore the effect of parameter estimation tend to be conservative. This problem does not arise for the particular case of CSR tests, as the observed number of events, n, is a sufficient estimator of λ, and conditional on n testing CSR is equivalent to test if the point process is Poisson (see Section 2.2.1).

The main advantage of Monte Carlo tests is that we are not constrained to known distribution theory, as the researcher can choose any informative statistic to test the null hypothesis. When asymptotic distribution theory is available, Monte Carlo tests provide an exact alternative for small samples and allow us to check whether the asymptotic theory can be applied. Discrepancies between the results of the classical and Monte Carlo tests may imply that the classical test is based on inappropriate distributional assumptions.

2.2.3 Properties of spatial point processes

In this section we introduce some properties that are commonly assumed in the analysis of spatial point processes. Let $N(W) = \#(\mathbf{x_i} \in W)$ be the random variable which measures the number of events observed in the planar region W, then

- A spatial point process is *stationary* if for any set of regions, $\{W_i, i = 1, \ldots, k\}$, the joint distribution of $N(W_1), N(W_2), \ldots, N(W_k)$ is invariant under translation of the regions by any set x.
- A spatial point process is *isotropic* if for any set of regions, $\{W_i, i = 1, \ldots, k\}$, the joint distribution of $N(W_1), N(W_2), \ldots N(W_k)$ is invariant under rotation of the union of the regions.

- A spatial point process is *orderly* (Diggle, 2003) when it does not contain duplicated events, i.e. if $N(x)$ is the number of events observed in the region $dx.$ and $|dx|$ its area

$$\lim_{|dx|\to 0} \frac{P\left(N(dx)\right) > 1}{|dx|} = 0$$

  This property is also referred as *simplicity* (Illian et al., 2008).

- A spatial point process is *second-order orderly* if for each pair of points x and y

$$\lim_{|dx|,|dy|\to 0} \frac{\left(P\left(N(dx)\right) > 1\right)\left(P\left(N(dy)\right) > 1\right)}{|dx||dy|} = 0$$

Stationarity and isotropy mean that the properties of the spatial point process are invariant under translation and rotation, respectively. Diggle (2003) argued that these properties are less restrictive that it might seem at first glance, and do not rule out the modeling of random heterogeneity. For instance, under these assumptions the heterogeneity observed in Figure 2.1 (centre) shall be interpreted as interaction between events. However, in the application to real data these assumption may be quite restrictive and unrealistic.

2.2.4 First and second-order characteristics

2.2.4.1 First-order intensity

The *first-order intensity function* of a point process describes its mean structure, and is key in dictating the spatial structure of its events. For this reason, obtaining accurate estimator of the first-order intensity has been a main issue in the analysis of spatial point processes. Formally, the first-order intensity function, $\lambda(\cdot)$ is defined as follows (Diggle, 2003)

$$\lambda(x) = \lim_{|dx|\to 0} \left\{ \frac{E\left[N(dx)\right]}{|dx|} \right\} \tag{2.2.1}$$

where E denotes expectation of a random variable, $|dx|$ and $N(dx)$ denote the area and the number of events of $\mathbf{X}$ observed in dx, respectively. Intuitively, $\lambda(x)|dx|$ is the probability for dx to contain exactly one event of the point process. A point process is *homogeneous* if its first-order intensity is constant, $\lambda(x) = \lambda > 0$, and *inhomogeneous* otherwise, i.e. when the probability of observing an event at x depends on the spatial location.

2.2.4.2 Second-order characteristics

Second-order characteristics analyze the relationship between pairs of events and are key to understand the nature of covariance structure in any observed pattern. For this reason, an important body of research in spatial point processes has focused on estimating second-order summary statistics. Below we introduce the more relevant second-order characteristics, and discuss their properties for stationary and isotropic point processes.

The *second-order intensity* function is defined as

$$\lambda_2(x,y) = \lim_{|dx|,|dy|\to 0} \left\{ \frac{E\left[N(dx)N(dy)\right]}{|dx||dy|} \right\} \tag{2.2.2}$$

$\lambda_2(x,y)|dx||dy|$ can be interpreted as the probability of observing exactly an event on $|dx|$ and one event on $|dy|$.

Given the first and second-order intensities we can define the *conditional intensity* function $\lambda_c(x|y) = \lambda_2(x,y)/\lambda(y)$, which measures the intensity at point x given that there is an event in y.

For a stationary point process $\lambda(x,y) \equiv \lambda(x-y)$, for a stationary and isotropic point process $\lambda(x,y) \equiv \lambda(\parallel x-y \parallel) = \lambda_2(r)$, where $r = \parallel x-y \parallel$. The ratio $\rho(r) = \lambda_2(r)/\lambda^2$ is known as *pair correlation function*, even this is not a correlation function in the statistical sense. We can also define the *covariance density* as $\gamma(r) = \lambda_2(r) - \lambda^2 = \lambda^2\left(\rho(r) - 1\right)$.

A particularly useful second-order statistic is the *reduced second moment measure* or *K-function* (Ripley, 1977), which characterizes the observed pattern in terms of counts of events within distance r of any arbitrary, i.e. randomly selected, event. The K-function is defined as

$$K\left(r\right) = \lambda^{-1} E\left[N_0(r)\right] \tag{2.2.3}$$

where $N_0(r)$ is the number of further events within distance r of an arbitrary event.

The K-function is invariant under random thinning, i.e. if each event of the point process is retained or not according to a series of mutually independent Bernouilli trials, the K-function of the resulting thinned process is equal to that of the original point process. The interpretation of the K-function is more intuitive than the second-order intensity

function. Furthermore, its definition as the mean of an observable value suggests a direct way to estimate the K-functions of any observed pattern. Let $\mathbf{X}$ be a spatial point pattern of size n observed on W, its K-function can be estimated as follows

$$\hat{K}(r) = \frac{|W|}{n(n-1)} \sum_{\mathbf{x_i},\mathbf{x_j} \in \mathbf{X} \bigcap W}^{i \neq j} I\left(\| \mathbf{x_i} - \mathbf{x_j} \| \leq r\right) w_{i,j}^{-1};\ 0 \leq r \leq r_{max} \tag{2.2.4}$$

where $I(\cdot)$ is the indicator function, $n/|W|$ is the empirical estimator of the first-order intensity, $w_{i,j}$ is an edge-correction term, and r_{max} is an upper bound for the distances at which the function can be evaluated. Both edge-correction and the maximum distance are necessary because the point pattern is observed in a bounded region. Edge or boundary effects are a well-known problem in spatial statistics, which arise when data are observed in a subset, W, of the region where the underlying process is defined. The unobserved events outside W may interact with observed events within W and affect the estimation of desirable quantities, but as those events remain unobserved, it is difficult to take proper account of these interactions. Edge-correction is a reweighted step developed to compensate this loss of information. The edge-corrector introduced by Ripley (1977) for the K-function estimator, $w_{i,j}$, is defined as the reciprocal of the proportion of the disc centered at location $\mathbf{x_i}$ with radius $\| \mathbf{x_i} - \mathbf{x_j} \|$ that lays inside W (see alternative edge-correctors in Baddeley and Turner (2005) and references therein).

Ripley (1977) showed that (2.2.4) is approximately unbiased for small values of r. In addition, due to edge-effects we need to determine a maximum distance, r_{max}, for $\hat{K}(r)$. as we cannot count the number of events at distance r of an observed event if the intersection between the disc with center x and radius r and W is the empty space. Thus r_{max} can de defined as the maximum distance for which any event in W can find at least another event within W at distance r_{max}.

In order to establish a link between $K(r)$ and $\lambda_2(r)$ we assume that the point process is orderly, which implies that multiple coincident events cannot occur, and second-order orderly (see Section 2.2.3). These properties mean that

$$\lim_{|dx| \to 0} \frac{E\left[N(dx)\right]}{P\left(N(dx) = 1\right)} = 1$$

and

$$\lim_{|dx|,|dy| \to 0} \frac{E\left[N(dx)N(dy)\right]}{P\left(N(dx) = N(dy) = 1\right)} = 1$$

Under these assumptions, the expected number of further events within distance r of an arbitrary event can be computed by integrating the second-order intensity over the disc with center the origin and radius r. Thus

$$\lambda K(r) = \int_0^{2\pi} \int_0^r \lambda_c(x|0) x dx d\theta$$

given that $\lambda_c(x|0) = \lambda_2(r)/\lambda^2$, we obtain the following expression

$$K(r) = 2\pi\lambda^{-2} \int_0^r \lambda_2(x) x dx \tag{2.2.5}$$

or conversely, $\lambda_2(r) = \lambda^2(2\pi r)^{-1}K'(r)$ and

$$\rho(r) = (2\pi r)^{-1}K'(r) \tag{2.2.6}$$

Under CSR $\lambda_2(r) = \lambda^2$ and, consequently $\rho(r) = 1$. If $\rho(r) > 1$, events at distance r are more frequent than expected for a completely random point process, which is typical of clustered or aggregated point processes. If $\rho(r) < 1$ events at distance r are less frequent than expected under CSR, i.e. we have a regular point process. Expression (2.2.5) implies that the K-function of a completely random point process is $K(r) = \pi r^2$. Hence, $K(r) > \pi r^2$ for clustered point processes, and $K(r) < \pi r^2$ for regular point processes. This property suggests that the distance between the empirical K-function of an observed pattern and πr^2 can be used as discrepancy measure to test for CSR. Figure 2.2 displays the estimated, (2.2.4), and the theoretical K-functions for the point patterns in Figure 2.1. For the completely random point pattern (left) both lines are coincident. As expected, the estimated K-function of the clustered point pattern is above the theoretical line. Finally, the estimated K-function for the regular point pattern is slightly lower than the theoretical line for small distances, it should be noted that the inhibition radius used to simulate this pattern was 0.03.

In the application to real data the K-function has some advantages over $\lambda_2(r)$ or $\rho(r)$, specially in small samples. Essentially both statistics are related to the distribution and the probability density functions of the distance between pairs of events in a point pattern, and the K-function can be estimated without the need of deciding how much to smooth the corresponding empirical distribution. However, from the mathematical viewpoint it may be more convenient to work with $\lambda_2(r)$ or $\rho(r)$ rather than with $K(r)$. In addition, if the dataset is large enough, we can easily obtain a histogram-type estimator of $\lambda_2(r)$ and $\rho(r)$, which is considered easier to interpret (Diggle, 2003).

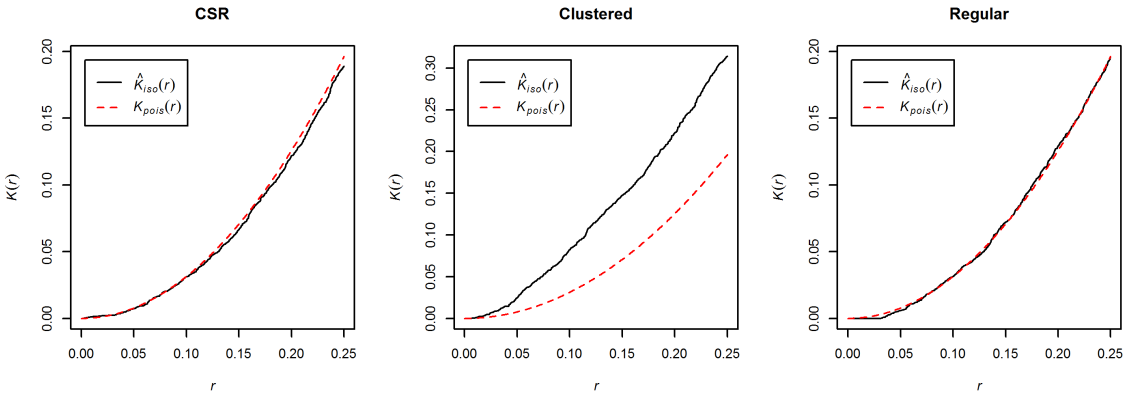


FIGURE 2.2: Estimated K-function for the point patterns in Figure 2.1. $\hat{K}_{iso}(r)$ estimated K-function with Ripley´s edge-corrector, $\hat{K}_{pois}(r) = \pi r^2$, that is the K-function under CSR

Given that in practice the stationarity assumption can be very restrictive and unrealistic, Baddeley et al. (2000) introduced *second-order reweighted stationary (SORS)* point processes, which first-order intensity is inhomogeneous and bounded away from 0 and

$$\rho(r) = \lambda_2(r)/\lambda(x)\lambda(y) \tag{2.2.7}$$

depends only on the distance $r = \| x - y \|$, in this case $\rho(r)$ is also named *pair correlation function*. This property is analogous to the assumption commonly made in the analysis of real-valued spatial processes that the mean value varies spatially but the variation around the local mean is stationary.

Baddeley et al. (2000) extended the K-function to the case of SORS point processes and defined the *inhomogeneous K-function* as follows

$$K_{inhom}(r) = \frac{1}{|W|} E\left[\sum_{\mathbf{x_i},\mathbf{x_j} \in \mathbf{X} \bigcap W}^{i \neq j} \frac{I\left(\| \mathbf{x_i} - \mathbf{x_j} \| \right)}{\lambda(\mathbf{x_i})\lambda(\mathbf{x_j})}\right] \tag{2.2.8}$$

The natural empirical estimator of the inhomogeneous K-function is

$$\hat{K}_{inhom}(r) = \frac{1}{|W|} \sum_{\mathbf{x_i},\mathbf{x_j} \in \mathbf{X} \bigcap W}^{i \neq j} \frac{I\left(\| \mathbf{x_i} - \mathbf{x_j} \| \leq r\right)}{\hat{\lambda}\left(\mathbf{x_i}\right)\hat{\lambda}\left(\mathbf{x_j}\right)} w_{i,j}^{-1};\ 0 \leq r \leq r_{max} \tag{2.2.9}$$

where $\hat{\lambda}\left(x\right)$ is the first-order intensity estimator (see Chapter 2 for details on first-order intensity estimation), $w_{i,j}$ is Ripley´s edge-corrector, and r_{max} is the maximum distances

at which the function can be evaluated.

The inhomogeneous K-function is linked to the pair correlation function (2.2.7) through the following expression

$$K_{inhom}(r) = 2\pi \int_0^r \rho(x)xdx \tag{2.2.10}$$

For a stationary point process $\lambda(x) = \lambda$, hence expression (2.2.8) reduces to (2.2.3) and (2.2.10) reduces to (2.2.5). For an inhomogeneous Poisson process (IPP), i.e. for a point process with non-constant intensity function but independent events, $K_{inhom}(r) = \pi r^2$

A transformation of the K-function commonly used in practice is the *L-function* introduced by Besag (1977), which for a stationary point process is defined as

$$L(r) = \sqrt{K(r)/\pi} \tag{2.2.11}$$

we can also define the *inhomogeneous L-function* as $L_{inhom}(r) = \sqrt{K_{inhom}(r)/\pi}$. The popularity of the L-function relies on the fact that under CSR $L(r) = r$, and for an inhomogeneous Poisson point process $L_{inhom}(r) = r$. It has been argued that deviations from a straight line are easier to detect and interpret than deviations from the curve πr^2, and, on the other hand, $\hat{L}(r)$ suffers a lower increase in variability for large values of r than $\hat{K}(r)$.

2.2.5 Empty space and nearest neighbor distributions

The K-function is a useful tool in the analysis of spatial point processes. However, as well as the first or second-order moments do not characterize the distribution of a random variable, the K-function does not characterize the distribution of a point process (Baddeley and Silverman, 1984). Higher order moments can be defined in terms of the joint intensity function for the occurrence of configurations of three, four or more events. However, as these functions depend on an increasing number of arguments, their interpretation may be difficult even for stationary point processes.

For these reasons, two additional summary descriptors that can also be used to test for CSR were defined (Diggle, 2003). These are the *nearest neighbor distribution function*, $G(u)$, and the *empty space distribution function*, $F(v)$. The former is defined as

$G(u) = P(u_i < u)$, where u_i is the distance from the ith event to the nearest event. Given m points in W, we define $F(v) = P(v_j < v)$, where v_j is the distance from the jth point to the nearest event.

The theoretical nearest neighbor and empty space distribution functions are intractable, except for homogeneous Poisson processes, which limits their use in point process modeling. However, both $G(u)$ and $F(v)$ can be used in goodness-of fit tests and to distinguish between models that cannot be discriminated comparing their second-order properties. In particular, the theoretical distributions of $G(u)$ and $F(v)$ under CSR depend exclusively on the number of events, n, and the region W. If we overlook edge-effects, the probability of finding an event within distance r from an arbitrary event is $\pi r^2|W|^{-1}$. Thus, since the event locations are mutually independent, the approximate nearest neighbor distribution is

$$G(u) = 1 - \left(1 - \pi u^2 |W|^{-1}\right)^{-1}$$

Taking into account that $\lambda = n|W|^{-1}$, a further approximation for large values of n is

$$G(u) = 1 - \exp\left(-\lambda \pi u^2\right);\ u > 0 \tag{2.2.12}$$

Similarly we obtain that under CSR the empty space distribution function, $F(v)$, is approximately

$$F(v) = 1 - \exp\left(-\lambda \pi v^2\right);\ v > 0 \tag{2.2.13}$$

i.e. for a completely random point process both distributions are equal.

2.2.6 Tests of complete spatial randomness

As indicated in Section 2.2.1, testing complete spatial randomness is the natural starting point in the analysis of any observed pattern, given that rejecting CSR is a minimum prerequisite to model any observed pattern. These tests are commonly used as an exploratory tool that provide information to formulate the alternative model, given that CSR acts as dividing hypothesis between regular and aggregated patterns.

From the pedagogic viewpoint, the CSR tests highlight the value of graphical methods, which may make formal testing unnecessary, the important role of Monte Carlo tests,

and the need of taking into account the dependence between multiple measures derived form the same point process.

In the literature we can find *quadrat counts tests*, which are based on partitions of the observation region, W; and Monte Carlo tests based on distance distributions, such as the nearest-neighbor and empty-space distribution functions (see Diggle (2003) and Illian et al. (2008) for further information). However, the most popular tools to test the CSR hypothesis are the Monte Carlo test based on the K-function and on the L-function, which can be easily extended to test for independence in inhomogeneous point processes, and to the multivariate case.

2.2.6.1 K-test for homogeneous point processes

We have seen in Section 2.2.4 that for a completely random point pattern $K(r) = \pi r^2$. Furthermore, comparison between the K-function of an observed pattern and its value under CSR yields

$$\begin{aligned} K(r) > \pi r^2 &\Rightarrow \text{ aggregated point process} \\ K(r) < \pi r^2 &\Rightarrow \text{regular point process} \end{aligned} \tag{2.2.14}$$

This relationship suggests using the K-function as statistic to test the hypothesis of CSR. Let $\mathbf{X}$ be a spatial point pattern of size n observed on W, the Monte Carlo CSR-test is implemented as follows: (i) obtain $\hat{K}_1(r)$ the estimated K-function for the observed pattern using (2.2.4); (ii) simulate $B-1$ homogeneous spatial Poisson point processes with n events and estimate the K-functions of each simulated pattern $\hat{K}_2(r), \ldots, \hat{K}_B(r)$; (iii) compute the upper and lower envelopes of the simulations, which are defined as

$$U(r) = \max_{j=2,\ldots,B} K_j(r)\,;\; L(r) = \min_{j=2,\ldots,B} K_j(r)$$

and (iv) completely random, aggregated or regular patterns are identified, depending on whether the K-function of the observed pattern is enclosed within the envelopes, higher than the upper envelope, or lower than the lower envelope, respectively. If the null hypothesis is rejected, the maximum r for which the K-function of the observed pattern is outside the confidence band determined by the envelopes indicates the radius of interaction between events.

It may be preferable using an equivalent test based on the L-function (2.2.11) for two reasons: (i) the fact that under CSR $L(r) = r$ provides a more intuitive graphical test; and (ii) the variability increase in the empirical estimators for large values of r, which reduces the power of test, is lower for the L-function.

Figure 2.3 shows the L-tests for the point patterns in Figure 2.1. These plots provide an intuitive interpretation of the L-tests. As expected for the homogeneous point process (left) the observed L-function lays within the envelopes, for the clustered point pattern (center) $\hat{L}(r)$ is above the upper envelope, and for the regular point pattern (right) $\hat{L}(r)$ is below the lower envelope for $r < 0.05$, i.e. the test was able to identify as significant slight differences between the estimated and the theoretical L-functions (see Figure 2.2) and to determine the radius of interaction. This example illustrates the power of this Monte Carlo test even for small point patterns ($n \approx 100$).

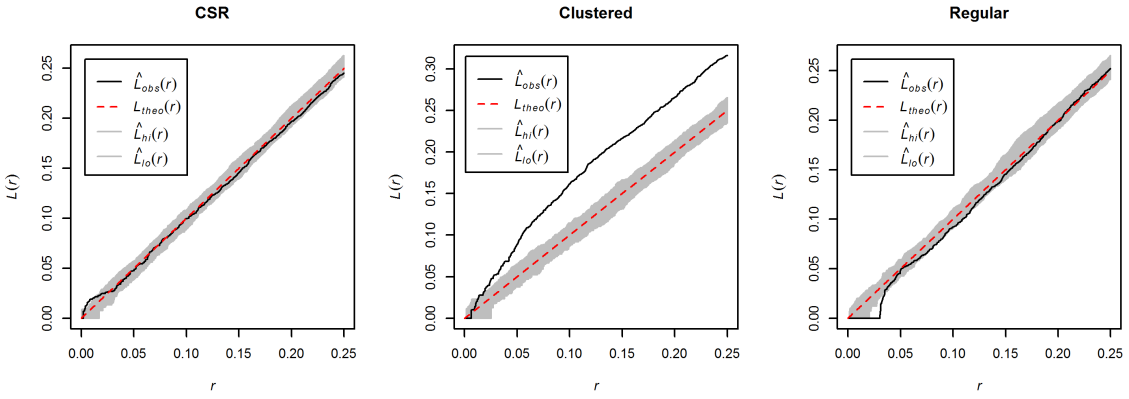


FIGURE 2.3: L-test for the spatial point patterns in Figure 2.1. $\hat{L}_{obs}(r)$ is the estimated L-function with Ripley´s edge-corrector, $\hat{L}_{theo}(r) = r$ is the K-function under CSR, $\hat{L}_{hi}(r)$ and $\hat{L}_{lo}(r)$ are the upper and lower envelopes of $B = 100$ simulations of the null hypothesis.

It should be noted that this CSR test is build on the stationarity assumption, i.e. we assume that the process is homogeneous and test for independence between events. However, if the first-order intensity of the observed pattern is not constant, the homogeneous K-function overstates the departure from CSR, and may lead to identify as interaction between events the inhomogeneity. This drawback highlights the need of testing the stationarity assumption prior to conduct the CSR-test, and the need of CSR-test for inhomogeneous point processes.

2.2.6.2 Testing CSR in inhomogeneous point processes

In practice, the stationarity assumption may be very restrictive and lead to a wrong characterization of the dependence structure of a spatial pattern. To overcome this problem the CSR test should be divided in two steps: (i) testing if the point process is stationary, and (ii) testing for independence between events. If the point process is stationary we are in the conditions to apply the K-test proposed above. When the stationarity assumption is rejected, CSR reduces to assume that the events are independent, i.e. under the null hypothesis we have an *inhomogeneous Poisson point process (IPP)*, which is defined by the following properties

IPP1 the expected number of events in any planar region, W, with surface area $|W|$ has Poisson distribution with mean $\lambda|W|$, where $\lambda = \int_W \lambda(x)dx$

IPP2 given n events in the region W, the events are an independent and random sample of the distribution on W with probability density function (pdf) proportional to $\lambda(x)$.

Inhomogeneous Poisson point processes provide a framework to introduce covariates in the analysis of spatial point processes. For instance, the spatial distribution of wildfires in a region may depend on the vegetation type, meteorological factors, and/or some spatial-varying socio-economic characteristics.

Stationarity test The stationarity assumption can be tested using quadrat counts (Hering et al., 2009), or measuring the discrepancy between the estimators of the homogeneous and inhomogeneous K-functions (Calduch, 2004). However, as the stationarity assumption implies that the first-order intensity is constant, stationary tests should be build on the first-order rather than on second-order properties. Comas et al. (2009) proposed a Monte Carlo test based on the discrepancy between the first-order intensity, $\lambda(x)$, and the constant intensity, $\lambda = \int_W \lambda(x)dx$ of the point process

$$\hat{S} = \int_W \parallel \hat{\lambda} - \hat{\lambda}(x) \parallel dx \tag{2.2.15}$$

where $\hat{\lambda}$ and $\hat{\lambda}(x)$ are estimators of the constant and local-varying intensities, respectively. The constant intensity can be estimated as $\hat{\lambda} = n/|W|$ or equivalently as $\hat{\lambda} = \int_W \hat{\lambda}(x)dx$. The stationarity assumption is accepted when $\hat{S} = 0$, while $\hat{S} > 0$ indicates that the point process is non-stationary. The Monte Carlo test is conducted comparing the discrepancy measure for the observed pattern, $\hat{S}_1$, with the discrepancies,

$\hat{S}_i, i = 2 \ldots, B$, obtained from simulated homogeneous spatial Poisson processes involving the same number of events as the observer pattern.

Inhomogeneous K-test Given that for an inhomogeneous Poisson point process $K_{inhom}(r) = \pi r^2$, the CSR K-test can be easily extend to test the Poisson assumption in inhomogeneous point processes. As explained in Section 2.2.6.1, the Monte Carlo test compares the estimated K-function of the observed pattern $\hat{K}_{inhom,1}(r)$, with the envelopes obtained from $B - 1$ simulations of inhomogeneous Poisson processes with first-order intensity $\hat{\lambda}(x)$

$$U(r) = \max_{j=2,\ldots,B} \hat{K}_{inhom,j}(r); \; L(r) = \min_{j=2,\ldots,B} \hat{K}_{inhom,j}(r)$$

when $\hat{K}_{inhom,1}(r)$ is enclosed in the confidence band determined by $U(r)$ and $L(r)$ the point process is Poisson. Values of $\hat{K}_{inhom,1}(r)$ above the upper envelope or below the lower envelope indicate clustered and regular patterns, respectively. As well as in the homogeneous case we can define an equivalent test based on the inhomogeneous L-function, $L_{inhom}(r) = \sqrt{K_{inhom}(r)/\pi}$.

Figure 2.4 shows a realization of an inhomogeneous Poisson point process with first-order intensity $\lambda(x, y) = 300 * \exp(-3x)$ observed on the unit square, and the results of the homogeneous and inhomogeneous L-tests applied to the simulated pattern. This example shows that, as explained above, the homogeneous L-test leads to interpret the inhomogeneity as interaction between events and identify the point pattern as clustered, while the inhomogeneous L-test does not detect any departure from the Poisson assumption.

2.3 Multitype point processes

A *multivariate* or *multitype point process* is a marked point process with categorical marks that define different groups. In this section we introduce the first and second-order characteristics and the main properties of multitype point processes .

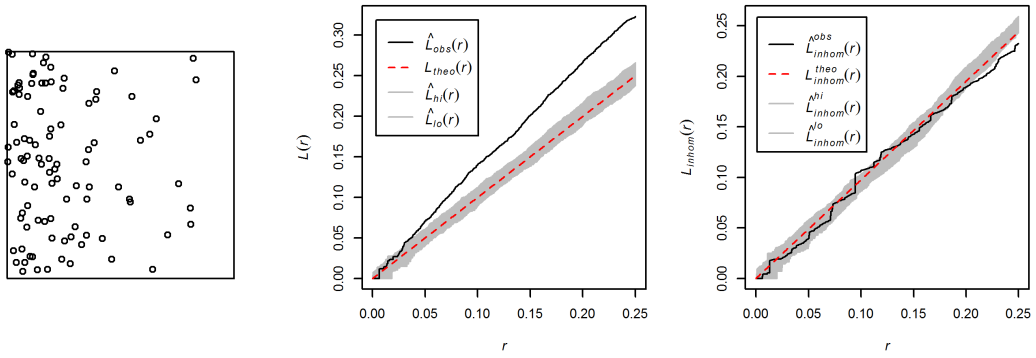


FIGURE 2.4: Simulated inhomogeneous spatial Poisson process (left) and comparison between the homogeneous (center) and inhomogeneous (right) L-tests.

2.3.1 First and second-order characteristics of multitype point processes

The first and second-order characteristics of multitype point processes are a natural extension of the univariate case. The first-order intensity of type j events is

$$\lambda_j(x) = \lim_{|dx|\to 0} \left\{ \frac{E\left[N_j(dx)\right]}{|dx|} \right\} \tag{2.3.1}$$

where $N_j(dx)$ is the number of type j events of $\mathbf{X}$ observed in dx. In a stationary point process, the first-order intensity of each type of events is constant, $\lambda_j(x) = \lambda_j = E\left[N_j(W)\right]/|W|$, and can be interpreted as the expected number of type j events per unit area.

Similarly, we can extend the second-order intensity function to the multivariate framework

$$\lambda_{ij}(x,y) = \lim_{|dx|,|dy|\to 0} \left\{ \frac{E\left[N_i(dx)N_j(dy)\right]}{|dx||dy|} \right\} \tag{2.3.2}$$

Intuitively, $\lambda_{ij}(x,y)|dx||dy|$ can be interpreted as the probability of observing exactly one type i event on $|dx|$ and one type j event on $|dy|$. As well as in the univariate case, $\lambda_{ij}(x,y) \equiv \lambda_{ij}(x-y)$ for a stationary point process, and $\lambda_{ij}(x,y) \equiv \lambda_{ij}(\parallel x-y \parallel) = \lambda_{ij}(r)$ if the point process is stationary and isotropic. Note that $\lambda_{ij}(r) = \lambda_{ji}(r)$.

Ripley (1981) extended the K-function to the multivariate case and introduced the *K-cross function*, which for a stationary and isotropic point processes is

$$K_{ij}(r) = \lambda^{-1} E\left[N_{0ij}(r)\right] \tag{2.3.3}$$

where $N_{0ij}(r)$ is the expected number of type j events within distance r of an arbitrary type i event. As well as in the univariate case the definition of the K-cross suggests a direct way to estimate this function

$$\hat{K}_{ij}(r) = \frac{1}{\lambda_i \lambda_j |W|} \sum_{\mathbf{x_k} \in \mathbf{X_i} \bigcap W} \sum_{\mathbf{x_l} \in \mathbf{X_j} \bigcap W} I\left(\| \mathbf{x_k} - \mathbf{x_l} \| \leq r\right) w_{k,l}^{-1};\ 0 \leq r \leq r_{max} \tag{2.3.4}$$

where $w_{k,l}$ is Ripley´s edge-corrector, r_{max} the maximum distance at which the K-cross is estimated, and $\hat{\lambda}_i$, $\hat{\lambda}_j$ are the empirical estimators of the intensity for each marginal process.

A similar argument to that used to obtain expression (2.2.5) provides a link between the K-cross and the bivariate second-order intensity

$$K_{ij}(r) = 2\pi \left(\lambda_i \lambda_j\right)^{-1} \int_0^r \lambda_{ij}(x) x dx \tag{2.3.5}$$

and, consequently we have that $K_{ij}(r) = K_{ji}(r)$.

For non-stationary point processes we can define the *inhomogeneous K-cross* as a natural extension of the inhomogeneous K-function (2.2.8), and its empirical estimator as

$$\hat{K}_{inhom,ij}(r) = \frac{1}{|W|} \sum_{\mathbf{x_k} \in \mathbf{X_i} \bigcap W} \sum_{\mathbf{x_l} \in \mathbf{X_j} \bigcap W} \frac{I\left(\| \mathbf{x_k} - \mathbf{x_l} \| \leq r\right)}{\lambda_i(x_k) \lambda_j(x_l)} w_{k,l}^{-1};\ 0 \leq r \leq r_{max} \tag{2.3.6}$$

where $\lambda_i(x)$ and $\lambda_j(x)$ are first-order intensities estimators for type i and type j point processes.

2.3.2 Independence and random labeling

To analyze the relationship between two types of events in a multitype point process we can consider two benchmark properties.

i *Independence*: the two types of events have been generated from two independent univariate point processes.

ii *Random labeling*: the two types of events have been generated by labeling the events of an univariate point process according to a series of mutually independent Bernouilli trials.

These hypothesis generate different K-cross functions. The K-cross function of a bivariate point process with independent type i and type j events is

$$K_{ij}(r) = \pi r^2$$

This follow from the fact that, if the two types of events are independent the expected number of type i events within distances r to an arbitrary type j event is $\lambda^2 \pi r^2$.

For any random labeled bivariate point process we have

$$K_{ii}(r) = K_{jj}(r) = K_{ij}(r) = K_r$$

where $K(r)$ is the K-function of the unlabeled univariate process. It should be noted that under random labeling the point processes with type i and type j events are random thinnings of the unlabeled point process including both type i and type j events. As the K-function is invariant under random thinning (Section 2.2.4), we have that $K_{ii}(r) = K_{jj}(r) = K(r)$. The same argument yields $K_{ij}(r) = K(r)$. Therefore independence and random labeling are equivalent if an only if both marginal processes are Poisson.

2.3.3 Tests of independence

When we are dealing with multitype point patterns, we may be interested in testing whether two types of events are independent. Spatial interaction between two types of events occurs when different types of events are either closer or further apart than expected under the assumption that the correspondent marginal processes are independent.

To test the hypothesis of independence between point patterns we can easily extend the univariate homogeneous and inhomogeneous K-tests to the multivariate framework. If both marginal processes are homogeneous the Monte Carlo test is conducted as follows: (i) use (2.3.4) to estimate the K-cross function of the observed pattern, $\hat{K}_{ij,1}(r)$; (ii) simulate $B-1$ bivariate point patterns with independent type i and type j events and estimate the K-cross of each simulated pattern $\hat{K}_{ij,2}(r), \ldots, \hat{K}_{ij,B}(r)$; To simulate

realizations of the null hypothesis, univariate point processes with intensities λ_i and λ_j are generated and labeled accordingly, these point patterns are combined to generate the bivariate point process; (iii) compute the upper, $U(r)$, and lower, $L(r)$, envelopes of the simulated patterns. Acceptance or rejection of independence depends on the position of $\hat{K}_{ij,1}(r)$ with respect to the envelopes, in particular

$$\begin{aligned} &\hat{K}_{ij,1}(r) > U(r) \Rightarrow \text{aggregation between type } i \text{ and type } j \text{ events} \\ &\hat{K}_{ij,1}(r) < L(r) \Rightarrow \text{inhibition between type } i \text{ and type } j \text{ events} \end{aligned} \tag{2.3.7}$$

If any marginal process is inhomogeneous, (2.3.6) is used to obtain $\hat{K}_{inhom,ij,1}$ in step (ii), and univariate point patterns with intensities $\hat{\lambda}_i(x)$ and $\hat{\lambda}_j(x)$ are generated to simulate the null hypothesis in step (iii).

For the same reasons argued in the univariate case, it is advisable using Monte Carlo tests based on the L-cross function, $L_{ij}(r) = \sqrt{K_{ij}(r)/\pi}$ or its inhomogeneous version, $L_{inhom,ij}(r))\sqrt{K_{inhom,ij}(r)/\pi}$.

2.3.4 The spatial relative risk function

A particular case of bivariate spatial point processes arises in environmental epidemiology, where researchers deal with datasets comprising the geographical location of disease cases, $\mathbf{X}$, and a random sample of the population at risk (commonly referred as controls), $\mathbf{Y}$, in a given region. A natural question in the analysis of such datasets is whether the disease risk varies spatially. In order to answer this question Bithell (1990) introduced the spatial relative risk function that describes the spatial variation in disease risk through the comparison of the spatial distribution of cases and controls. Let $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x}_{n_1}\}$ and $\mathbf{Y} = \{\mathbf{y_1}, \ldots, \mathbf{y}_{n_2}\}$ be realizations of spatial Poisson point processes with first-order intensities $\lambda_1(x)$ and $\lambda_2(x)$ observed in a bounded region $W \subset \mathbb{R}^2$, the *spatial relative risk function* is defined as

$$r(x) = \frac{\lambda_1(x)}{\lambda_2(x)}, \; x \in W \tag{2.3.8}$$

Kelsall and Diggle (1995a,b) recommended the use of the *log-relative risk function*, $\rho(x) = \log(r(x))$ in order to handle the densities of cases and controls symmetrically, as the number of controls can be much larger than the number of cases. Conditional on the number of events, n_1 and n_2, the spatial patterns of cases and controls can be treated as independent random samples from probability distributions with density functions $f(x)$ and $g(x)$ proportional to $\lambda_1(x)$, $\lambda_2(x)$. This duality was used by Kelsall and Diggle

(1995a) to develop the theoretical properties of the kernel estimator of $\rho(x)$. Thus, in advance we consider an alternative definition of the spatial relative risk function

$$r(x) = \frac{f(x)}{g(x)},\ x \in W \tag{2.3.9}$$

Kelsall and Diggle (1995a) used the kernel density estimators of f and g to estimate the log-relative risk function, $\rho(x) = \log\left(f(x)/g(x)\right)$, as

$$\hat{\rho}(x) = \log\left(\frac{\hat{f}_{h_f}(x)}{\hat{g}_{h_g}(x)}\right) = \log\left(\frac{\left(p_{h_f}(x)n_1\right)^{-1}\sum_{i=1}^{n_1} k_{h_f}\left(x-\mathbf{x_i}\right)}{\left(p_{h_g}(y)n_2\right)^{-1}\sum_{j=1}^{n_2} k_{h_g}\left(y-\mathbf{y_j}\right)}\right) \tag{2.3.10}$$

where the kernel $k(\cdot)$ is a bivariate radially symmetric density function, $p_{h_f}(x)$, $p_{h_g}(y)$ the respective edge-correction terms, and h_f, h_g bandwidth parameters. Although non-radially symmetric kernels or matrix bandwidths can be used in (2.3.10), in geographical epidemiology it is natural to work with isotropic kernels. The performance of $\hat{\rho}(x)$ can be measured in terms of its mean integrated square error:

$$MISE\left(h_f, h_g\right) = \int_W \left(\rho(x) - \hat{\rho}(x)\right)^2 dx$$

Overlooking the edge-corrector and assuming that f and g have bounded and continuous partial derivatives up to order 2, and $n_1, n_2 \to \infty$, $h_f, h_g \to 0$ such that $n_1 h_f^2, n_2 h_g^2 \to \infty$, Kelsall and Diggle (1995a) obtained the asymptotic expression of $MISE\left(h_f, h_g\right)$

$$\begin{aligned} AMISE\left(h_f, h_g\right) = & \frac{1}{4}\int_W \left(\frac{\mu_2(k)h_f^2\nabla^2 f(x)}{f(x)} - \frac{\mu_2(k)h_g^2\nabla^2 g(x)}{g(x)}\right)^2 dx \\ & + \int_W \left(\frac{R(k)f(x)^{-1}}{n_1 h_f^2} + \frac{R(k)g(x)^{-1}}{n_2 h_g^2}\right) \end{aligned} \tag{2.3.11}$$

Kelsall and Diggle (1995a) also developed a least-squares cross-validation bandwidth selector that defines the optimal bandwidths h_f, h_g as those minimizing

$$\begin{aligned} LSCV(h_f, h_g) = & -\int_{\mathbf{R}^2} \hat{\rho}(x)^2 dx - 2n_1^{-1}\sum_{i=1}^{n_1} \log\left(\frac{\hat{f}_{h_f}^{-i}(x_i)}{\hat{g}_{h_g}(x_i)}\right)\left(\hat{f}_{h_f}^{-i}(x_i)\right) \\ & + 2n_2^{-1}\sum_{j=1}^{n_2} \log\left(\frac{\hat{f}_{h_f}(y_j)}{\hat{g}_{h_g}^{-j}(y_j)}\right)\left(\hat{g}_{h_g}^{-j}(y_j)\right) \end{aligned} \tag{2.3.12}$$

and showed through a simulation study the convenience of using the same bandwidth ($h = h_f = h_g$) to estimate the case and control densities in (2.3.10) as it leads to a bias

cancellation in areas where $f = g$ and simplifies the data-driven procedure. Furthermore, considering the same bandwidth parameter for cases and controls we do not need to include the edge-correction terms in (2.3.10) as they cancelate.

2.4 Spatio-temporal point processes

A spatio-temporal point process is a stochastic process that generates events, $\mathbf{S} = \{(\mathbf{x_1}, \mathbf{t_1}), \ldots, (\mathbf{x_N}, \mathbf{t_N})\}$, on $W \times T \subset \mathbb{R}^2 \times \mathbb{R}^+$ and each event is represented by two spatial coordinates, $\mathbf{x_i} = (\mathbf{x_{i1}}, \mathbf{x_{i2}})$, and one temporal coordinate, $\mathbf{t_i}$. Given the different nature of the temporal component spatio-temporal point processes cannot be treated as point processes in $\mathbb{R}^3$. In the spatial setting, we assume that the space where events occur is a continuous region of the plane. In the extension to the spatio-temporal framework it can be useful to allow either the spatial or temporal dimension to be discrete (see further discussion and examples in Diggle (2013)). Spatially discrete and temporally discrete point patterns can formally be treated as multivariate temporal and spatial point processes, respectively.

This section focuses on the analysis of *continuous spatio-temporal point processes*, i.e. point processes with continuous spatial component and continuous and orderly temporal component. Section 2.4.1 introduces the first and second-order characteristics of spatio-temporal point processes, Sections 2.4.2 and 2.4.3 discuss first-order separability and the separability tests currently available. Finally Section 2.4.4 introduces the spatio-temporal relative risk function and its kernel estimator.

2.4.1 First and second-order characteristics

The first-order intensity function of a spatial point process (2.2.1) can be extended to the spatio-temporal framework to define the *spatio-temporal intensity function* (STIF) (Diggle, 2013) as

$$\lambda(x,t) = \lim_{|dx \times dt| \to 0} \left\{ \frac{E\left[N(dx,dt)\right]}{|dx \times dt|} \right\} \tag{2.4.1}$$

where $N(dx,dt)$ represents the number of events in the volume $dx \times dt$, dx is an infinitesimal disc containing the location x, and dt is an infinitesimal interval containing t.

For a spatially or temporally stationary process $\lambda(x,t)$ is independent of x or t respectively. For a spatio-temporally stationary, i.e. homogeneous, point process the intensity function assumes a constant value, $\lambda(x,t) = \lambda$. Intuitively, λ represents the expected number of events per unit area and unit time. From a strict mathematical perspective it makes no sense to talk about the spatial or temporal marginal characteristics of a spatio-temporal point process, as both the mean number of events per unit area and per unit time are infinite. In practice we only observe a spatio-temporal point process on a finite region $W \times T$, thus the spatial and temporal intensity functions can be defined as

$$\lambda_1(x) = \int_T \lambda(x,t)dt, \quad \lambda_2(t) = \int_W \lambda(x,t)dx \tag{2.4.2}$$

Given that point process intensities are not normalized, the spatio-temporal intensity function $\lambda(x,t)$ can be interpreted in three different ways: as a joint spatio-temporal intensity; as a spatial intensity conditional to a given value of t, or as a temporal intensity conditional to a given value of x.

The history of a spatio-temporal point process at time t, $H_t = \{(\mathbf{x_i}, \mathbf{t_i})\,; \mathbf{t_i} < t\}$, is the collection of all events of the process that occur before time t. The *conditional intensity function* (CIF) of a spatio-temporal point process, $\lambda(x,t|H_t)$, is the spatial first-order intensity at time t conditional on the history, H_t, i.e. on the process up to time t. More formally, if $N(dx,dt)$ is the number of events on an infinitesimal volume $dx \times dt$

$$\lambda(x,t|H_t) = \lim_{|dx \times dt| \to 0} \left\{ \frac{E\left[N(dx,dt)|H_t\right]}{|dx \times dt|} \right\} \tag{2.4.3}$$

Intuitively the conditional intensity describes how the likelihood of observing an event at location x and time t changes as a realization of the process up to, but not including, time t develops over time. While different point processes can share the same first-order intensity, the conditional intensity function characterizes an orderly spatio-temporal point process uniquely. Particularly, a *spatio-temporal Poisson point process* can be defined as a continuous and orderly process for which $\lambda\left(x,t|H_t\right) = \lambda(x,t)$, for all $(x,t) \in \mathbb{R}^2 \times \mathbb{R}^+$.

The *second-order spatio-temporal intensity function* is a natural extension of the second-order intensity of a spatial point process (2.2.2)

$$\lambda_2\left((x,t),(y,s)\right) = \lim_{|dx \times dt|, |dy \times ds| \to 0} \left\{ \frac{E\left[N(dx,dt)N(dy,ds)\right]}{|dx \times dt||dy \times ds|} \right\} \tag{2.4.4}$$

$\lambda_2((x,t),(y,s))|dx \times dt||dy \times ds|$ can be interpreted as the probability of observing exactly an event on $|dx \times dt|$ and one event on $|dy \times ds|$. The *second-order conditional spatio-temporal intensity function*, $\lambda_c((x,t)|(y,s)) = \lambda_2((x,t),(y,s))/\lambda(y,s)$, measures the intensity at point (x,t) conditional on the information that there is an event in (y,s).

For a spatio-temporally stationary, isotropic point process $\lambda_2((x,t),(y,s))$ reduces to $\lambda_2(u,v)$, where $u = \| x - y \|$ and $v = \| t - s \|$, and we can define the *spatio-temporal pair correlation function* as $\rho(u,v) = \lambda_2(u,v)/\lambda^2$. Note that in the spatio-temporal framework *isotropic* is a shorthand for spatially isotropic and temporally reversible.

As in the spatial framework, the spatio-temporal stationarity assumption can be very restrictive and unrealistic when dealing with real data. For this reason Gabriel and Diggle (2009) extended *second-order reweighted stationarity* (Baddeley et al., 2000) to the spatio-temporal framework. An inhomogeneous spatio-temporal point process which intensity function is bounded away from 0 is *(second-order) intensity-reweighted stationary* and isotropic if its pair correlation function, $\rho(u,v) = \lambda_2(u,v)/\lambda(x,t)\lambda(y,s)$, depends only on the spatio-temporal difference vector (u,v).

The *spatio-temporal K-function* of a stationary, isotropic point process is defined as

$$K_{ST}(u,v) = \lambda^{-1} E[N_0(u,v)] \tag{2.4.5}$$

where $N_0(u,v)$ is the number of further events within distance u and time v of an arbitrary event. The K-function of a second-order reweighted stationary point process can be obtained as a natural extension of expression (2.2.8). The same arguments used in Section 2.2.4 to link the spatial pair correlation function and the K-function yield the following expression for the *spatio-temporal inhomogeneous K-function*

$$K_{ST}(u,v) = 2\pi \int_0^u \int_0^v \rho(u',v')u'du'dv' \tag{2.4.6}$$

For both homogeneous and inhomogeneous spatio-temporal Poisson point processes $K_{ST}(u,v) = \pi u^2 v; u,v \geq 0$. Values of $K_{ST}(u,v)$ above and below $\pi u^2 v$ indicate, respectively, spatio-temporal clustering and inhibition between events within spatial and temporal distances u and v. Moreover, separability of $K_{ST}(u,v)$ into purely spatial and temporal components, $K_{ST}(u,v) = K_S(u)K_T(v)$, indicates absence of spatio-temporal interaction (Diggle et al., 1995).

Gabriel and Diggle (2009) proposed the empirical estimator of the inhomogeneous spatio-temporal K-function

$$\hat{K}_{ST}(u,v) = \frac{1}{|W \times T|} \frac{n}{n_v} \sum_{i=1}^{n_v} \sum_{j>i} \frac{I\left(\| \mathbf{x_i} - \mathbf{x_j} \| \leq u\right) I\left(\| \mathbf{t_i} - \mathbf{t_j} \| \leq v\right)}{\hat{\lambda}\left(\mathbf{x_i}, \mathbf{t_i}\right) \hat{\lambda}\left(\mathbf{x_j}, \mathbf{t_j}\right)} w_{i,j}^{-1} \tag{2.4.7}$$

where $\hat{\lambda}\left(x,t\right)$ is an estimator of the first-order intensity function, $w_{i,j}$ is Ripley´s spatial edge-corrector (Ripley, 1977), n/n_v , where n_v denotes the number of events for which $t_i \leq v$, is the temporal edge-corrector.

We cannot distinguish between first and second-order effects in an observed spatio-temporal point pattern without any additional information. For this reason, a common practice is to assume that the first-order effects are separable, i.e. the spatio-temporal intensity function can be decomposed into the product of its spatial and temporal components, $\lambda(x,t) = \lambda_1(x)\lambda_2(t)$, and interpret any non-separable feature as second-order effects. Under this assumption Gabriel and Diggle (2009) introduced Monte Carlo tests based on the discrepancy between the observed and the theoretical K-functions to test the Poisson assumption and independence between the spatial and temporal components.

2.4.2 Separability

Estimating the spatio-temporal intensity function and the spatio-temporal conditional intensity function are main issues in the analysis of inhomogeneous spatio-temporal point processes. These characteristics allow to describe the distribution of events on the observation domain, and the intensity function is needed to estimate the second-order structure, as we can see in expression (2.4.7). However, modeling the joint distribution of spatial locations and times of occurrence can be a challenging task, difficulty increasing when marks and covariates are available. Most of the models commonly used to this purpose assume a product form for these functions, i.e. they assume that the point process is separable (Cressie, 1993). Bellow we introduce separability for the spatio-temporal intensity function, equivalent definitions can be provided for the conditional intensity (Díaz-Avalos et al., 2013; Schoenberg, 2004). A spatio-temporal point process, S, is *first-order separable* if:

$$\lambda(x,t) = \lambda_1(x)\lambda_2(t) \tag{2.4.8}$$

A spatio-temporal marked point process, $S = \{(\mathbf{x_i}, \mathbf{t_i}, \mathbf{m_i}), i = 1, \ldots, N\} \in \mathbb{R}^2 \times \mathbb{R}^+ \times \mathbb{R}^d$ is *completely separable* when its first-order intensity can be decomposed as follows

$$\lambda(x, t, m) = \lambda_1(x)\lambda_2(t)f(m) \tag{2.4.9}$$

When the point process depends on a set of covariates, $c_1(x, t), \ldots, c_k(x, t)$, complete separability implies that

$$\lambda(x, t, m, c_1(x, t), ..., c_k(x, t)) = \lambda_1(x)\lambda_2(t)f(m)g_1\left(c_1(t, x)\right)...g_k\left(c_k(t, x)\right) \tag{2.4.10}$$

Separability is a desirable condition, as the product form of the intensity function simplifies its estimation and avoids the curse of dimensionality, specially when several marks and covariates are involved in the model. However this assumption can be quite restrictive. For instance, a separable model for wildfires would assume that the spatial distribution of wildfire risk does not vary over time. Despite the importance of this assumption, few works have addressed a rigorous analysis of separability. Ogata (1988), and Schoenberg (2003) used parametric rescaling methods to observe departures from separability in the epidemic-type aftershock sequence (ETAS) model for earthquake occurrences, and Guttorp and Minin (1995) detected nonseparability in precipitation data from Middletown, Pennsylvania. Several authors have investigated spatio-temporal changes in mark distributions quite generally (e.g., Ogata and Katsura (1993), and Kagan (1999) in seismology; Johnson (1996), and Flannigan and Wotton (2001) in the case of wildfires), and separability tests have been constructed for time series and spatial autoregressive processes (e.g., Shitan and Brockwell (1995)).

Recently, some works developed nonparametric separability tests for the conditional intensity function of spatio-temporal point processes (Assuncáo and Maia, 2007; Chang and Schoenberg, 2011; Díaz-Avalos et al., 2013; Schoenberg, 2004). In the next section we discuss the separability tests introduced by Schoenberg (2004) and Díaz-Avalos et al. (2013).

2.4.3 Nonparametric separability tests

Schoenberg (2004) and Díaz-Avalos et al. (2013) developed Monte Carlo separability tests based on the comparison between the separable, $\lambda^S(x, t|H_t)$, and nonseparable, $\lambda^S(x, t|H_t)$, conditional intensity functions, which should be similar if the point process is separable. These tests have been developed for spatio-temporal marked point processes and for point processes that depend on covariates, but in this section we describe

the particular case of separability between the spatial and temporal components of unmarked spatio-temporal point processes.

Schoenberg (2004) and Díaz-Avalos et al. (2013) used kernel estimators of the spatio-temporal conditional intensity functions to conduct the nonparametric separability test. The kernel estimator of the nonseparable spatio-temporal intensity function

$$\hat{\lambda}^{NS}(x,t|H_t) = \int_{\mathbb{R}^2\times\mathbb{R}^+} k_3(u-x, v-t)dS(x,t) \tag{2.4.11}$$

for the separable spatio-temporal intensity function, we have

$$\hat{\lambda}^{S}(x,t|H_t) = \hat{\lambda}_1(x)\hat{\lambda}_2(t) \tag{2.4.12}$$

where

$$\hat{\lambda}_1(x) = \int_{\mathbb{R}^2\times\mathbb{R}^+} k_2(u-x)dS(x,t)$$

and

$$\hat{\lambda}_2(t) = \frac{1}{N}\int_{\mathbb{R}^2\times\mathbb{R}^+} k_1(v-t)dS(x,t)$$

are the kernel estimators of the marginal spatial and temporal intensities, respectively, N is the number of events of the observed spatio-temporal pattern, and $k_d(\cdot), d = 1, 2, 3$ are d-dimensional kernel functions.

It should be noted that (2.4.11) and (2.4.12) are kernel estimators of the first-order intensity rather than the conditional intensity function, consequently the discrepancy measures proposed by Schoenberg (2004) and Díaz-Avalos et al. (2013) can be viewed as tests on whether the overall intensity rather than the conditional intensity has a product form. Furthermore, as we cannot distinguish between hetereogeneity and interaction in an observed point pattern without additional information, we can assume that the spatio-temporal point process is Poisson, estimate its intensity function, and then estimate the second-order properties to test the Poisson assumption. Taking into account these considerations, in advance we focus on testing the separability of the spatio-temporal intensity function, which for a Poisson process fulfills $\lambda\,(x,t|H_t) = \lambda(x,t)$.

Schoenberg (2004) proposed separability tests based on the standardized maximum and minimum absolute distances:

$$S_1 = \sup_{(x,t)\in W\times T}\left[\frac{\hat{\lambda}^{NS}(x,t)-\hat{\lambda}^{S}(x,t)}{\hat{\lambda}^{S}(x,t)}\right] \tag{2.4.13}$$

$$S_2 = \inf_{(x,t)\in W\times T}\left[\frac{\hat{\lambda}^{NS}(x,t)-\hat{\lambda}^{S}(x,t)}{\hat{\lambda}^{S}(x,t)}\right] \tag{2.4.14}$$

a Cramer-Von-Mises type statistic:

$$S_3 = \int_0^T \int_{\mathbb{R}^2} \left(\hat{\lambda}^{NS}(x,t)-\hat{\lambda}^{S}(x,t)\right)^2 dxdt \tag{2.4.15}$$

and a log-likelihood separability test:

$$\begin{aligned} S_4 = &\int_X \left[\log\left(\hat{\lambda}^{NS}(x,t)\right)-\log\left(\hat{\lambda}^{S}(x,t)\right)\right] dN \\ &-\int_0^T \int_{\mathbb{R}^2} \left(\hat{\lambda}^{NS}(x,t)-\hat{\lambda}^{S}(x,t)\right) dxdt \end{aligned} \tag{2.4.16}$$

In order to introduce two new discrepancy measures to test the separability assumption, Díaz-Avalos et al. (2013) took into account that the ratio

$$f(x,t) = \frac{\lambda\,(x,t)}{\int_{W\times T}\lambda\,(x,t)\,dxdt}$$

is a density in $\mathbb{R}^2\times\mathbb{R}^+$. Conditional on the number of events, $N=n$, the separable and nonseparable estimators of $f(x,t)$ evaluated at the events of the observed pattern, $\mathbf{S}=\{(\mathbf{x_1},\mathbf{t_1}),\ldots,(\mathbf{x_n},\mathbf{t_n})\}$, are

$$p_i^S = \frac{\hat{\lambda}^{S}\left(\mathbf{x_i},\mathbf{t_i}\right)}{\sum_{i=1}^{n}\hat{\lambda}^{NS}\left(\mathbf{x_i},\mathbf{t_i}\right)}$$

and

$$p_i^{NS} = \frac{\hat{\lambda}^{NS}\left(\mathbf{x_i},\mathbf{t_i}\right)}{\sum_{i=1}^{n}\hat{\lambda}^{NS}\left(\mathbf{x_i},\mathbf{t_i}\right)}$$

The Kullback-Leibler (KL) measure and the Hellinger (H) distance measure the difference between the separable, p^S, and nonseparable, p^{NS}, ratios

$$KL = \sum_{i=1}^{n}\log\left(\frac{\hat{p}^{NS}}{\hat{p}^{S}}\right)\hat{p}^{NS} \tag{2.4.17}$$

and

$$H = \sum_{i=1}^{n} \left\{ \left(\sqrt{\hat{p}^{NS}} - \sqrt{\hat{p}^{S}} \right)^2 \right\} \tag{2.4.18}$$

and define two new statistics to test whether an observed pattern is separable.

Abnormally large values of (2.4.13 - 2.4.18) indicate departure from the separability assumption (2.4.8).

In a simulation study conducted to test the performance of tests (2.4.13)- (2.4.16) for separability between marks and the spatio-temporal coordinates, Schoenberg (2004) found that S_3 and S_4 are quite powerful, specially to detect gradual departure from separability. Díaz-Avalos et al. (2013) found that KL and H are competitive with the former tests. In particular, for both Poisson and clustered point processes, they found that when testing for separability between the spatial and temporal components, the probability of type II error for the Monte Carlo tests based on (2.4.17) and (2.4.18) decreased faster than for any other test statistic.

2.4.4 The spatio-temporal relative risk function

Sarojinie Fernando and Hazelton (2014) extended the spatial relative risk function introduced in Section 2.3.4 to the spatio-temporal framework in order to handle epidemiological data comprising the spatial location and time of occurrence of cases and controls, and provide tools that allow researchers to check whether spatial patterns of disease risk change over time. The *spatio-temporal relative risk function* is

$$r(x,t) = \frac{f(x,t)}{g(x,t)},\ x \in W, t \in T \tag{2.4.19}$$

where $t \in T$ is the time of occurrence of the event observed at location x, and $f(x,t)$, $g(x,t)$ denote the joint distribution of spatial locations and time of occurrence for cases and controls, respectively. As in the spatial case, it is advisable using the *spatio-temporal log-relative risk function*, $\rho(x,t) = \log r(x,t)$.

Conditional on the number of events, $N_j = n_j, j = 1,2$, let $\{(\mathbf{x_1}, \mathbf{t_1}), \ldots, (\mathbf{x_{n_1}}, \mathbf{t_{n_1}})\}$, $\{(\mathbf{x_{n_1+1}}, \mathbf{t_{n_1+1}}), \ldots, (\mathbf{x_n}, \mathbf{t_n})\}$ be the observed patterns of cases and controls, and $n = n_1 + n_2$ the total sample size. In order to extend the kernel estimator (2.3.10) proposed

by Kelsall and Diggle (1995a) to the spatio-temporal framework, we consider the kernel estimator of the spatio-temporal density for the case data.

$$\hat{f}_{h_s,h_t}(x,t) = \frac{1}{np_{h_s,h_t}(x,t)} \sum_{i=1}^{n_1} k_{s,h_s}\left(x - \mathbf{x_i}\right) k_{t,h_t}\left(t - \mathbf{t_i}\right) \tag{2.4.20}$$

where the kernel functions, $k_x(\cdot)$ and $k_t(\cdot)$, are a spherically symmetric bivariate density function and a univariate density function. h_s and h_t are the bandwidth parameters for the spatial and temporal component, $p_{h_s,h_t}(x,t) = \int_{W\times T} k_{s,h_s}(x-y)k_{t,h_t}(t-s)dyds = \int_W k_{s,H_s}(x-y)dy \int_T k_{t,h_t}(t-s)ds$ is the spatio-temporal edge-correction term, where $p_{h_s}(x) = \int_W k_{s,h_s}(x-y)dy$ and $p_{h_t}(t) = \int_T k_{t,h_t}(t-s)ds$ represent, respectively, the bivariate edge-corrector for the spatial locations and the univariate edge-correction for the time of occurrence. The kernel estimator for the joint density of control data, $g(x,t)$ is defined in the same way. Following the suggestion of Kelsall and Diggle (1995a) in the spatial framework, we can use the same bandwidth parameters in the kernel estimators of the spatio-temporal densities of cases and controls and define (Sarojinie Fernando and Hazelton, 2014)

$$\hat{\rho}(x,t) = \log\left(\frac{\hat{f}_{h_s,h_t}(x,t)}{\hat{g}_{h_s,h_t}(x,t)}\right) \tag{2.4.21}$$

As usual in kernel smoothing the exact properties of $\hat{\rho}(x,t)$ are intractable, but under some regularity conditions we can derive closed-form asymptotic properties. Assuming that: (i) $n \to \infty$, h_s, $h_t \to 0$ such that $nh_s^2, nh_t \to \infty$, and $n_1/n_2 \to c$, where $c > 0$ is a finite constant, and (ii) f and g are bounded away from 0 and fulfill the regularity conditions needed to develop the asymptotic properties of multivariate kernel density estimators (Wand and Jones, 1995), Sarojinie Fernando and Hazelton (2014) obtained the following expressions for the bias and variance of (2.4.19) in any interior point of $W \times T$

$$\begin{aligned} Bias\left(\hat{\rho}(x,t)\right) = {} & \frac{h_s^2\mu_2(k_s)}{2}\left[\frac{\nabla_x^2 f(x,t)}{f(x,t)} - \frac{\nabla_x^2 g(x,t)}{g(x,t)}\right] \\ & + \frac{h_t^2\mu_2(k_t)}{2}\left[\frac{\nabla_t^2 f(x,t)}{f(x,t)} - \frac{\nabla_t^2 g(x,t)}{g(x,t)}\right] + o\left(h_s^2 + h_t^2\right) \end{aligned} \tag{2.4.22}$$

$$Var\left(\hat{\rho}(x,t)\right) = \frac{R(k_s)R(k_t)}{h_s^2 h_t}\left[\frac{1}{n_1 f(x,t)} + \frac{1}{n_2 g(x,t)}\right] + o\left(n^{-1}h_s^{-2}h_t^{-1}\right] \tag{2.4.23}$$

Sarojinie Fernando and Hazelton (2014) also extended the least-squares cross-validation bandwidth selector for the spatial log-relative risk function (2.3.12), and defined the

optimal bandwidth vector $h = c(h_s, h_t)$ as that minimizing

$$LSCV(h_s, h_t) = -\int_{\mathbf{R}^2\times\mathbb{R}^+} \hat{\rho}(x,t)^2 dxdt - 2n_1^{-1}\sum_{i=1}^{n_1} \log\left(\frac{\hat{f}^{-i}_{h_s,h_t}(\mathbf{x_i},\mathbf{t_i})}{\hat{g}_{h_s,h_t}(\mathbf{x_i},\mathbf{t_i}}\right)\left(\hat{f}^{-i}_{h_s,h_t}(\mathbf{x_i},\mathbf{t_i})\right)^{-1} + 2n_2^{-1}\sum_{j=n_1+1}^{n} \log\left(\frac{\hat{f}_{h_s,h_t}(\mathbf{x_j},\mathbf{t_j})}{\hat{g}^{-j}_{h_s,h_t}(\mathbf{x_j},\mathbf{t_j})}\right)\left(\hat{g}^{-j}_{h_g}(\mathbf{x_j},\mathbf{t_j})\right)^{-1} \tag{2.4.24}$$

In many situations the control population remains unchanged through time, and the spatio-temporal relative risk function (2.4.19) reduces to

$$r(x) = \frac{f(x,t)}{g(x)},\ x \in W, t \in T \tag{2.4.25}$$

which can be estimated by

$$\hat{\rho}(x,t) = \log\left(\frac{\hat{f}_{h_s,h_t}(x,t)}{\hat{g}_{h_s}(x)}\right) \tag{2.4.26}$$

where, as in expression (2.3.10), $\hat{g}_{h_s}(x)$ is the kernel estimator of the bivariate density of controls.

Assuming the same regularity conditions as above, the bias and variance of (2.4.26) are

$$Bias\left(\hat{\rho}(x,t)\right) = \frac{1}{2}\left[\frac{h_s^2\mu_2(k_s)\nabla_x^2 f(x,t) + h_t^2\mu_2(k_t)\nabla_t^2 f(x,t)}{f(x,t)} - \frac{h_s^2\mu_2(k_s)\nabla^2 g(x)}{g(x)}\right] + o\left(h_s^2 + h_t^2\right) \tag{2.4.27}$$

and

$$Var\left(\hat{\rho}(x,t)\right) = \frac{R(k_s)R(k_t)f(x,t)^{-1}}{n_1 h_s^2 h_t} + \frac{R(k_s)g(x)^{-1}}{n_2 h_s^2} + o\left(n^{-1}h_s^{-2}\right) \tag{2.4.28}$$

Although, in comparison with the first term, the second term in expression (2.4.28) is asymptotically negligible, maintaining this term provides a better estimation of finite sampling behavior.

Finally, Sarojinie Fernando and Hazelton (2014) argued that the LSCV bandwidth selector cannot be extended to the context of spatio-temporal relative-risk with time-invariant control density. This extension may imply evaluate $f(x,t)$ at control data point, which is not possible because controls are not indexed by time.

Chapter 3

Consistent kernel estimator of the first-order intensity function

3.1 Introduction

Suppose that $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_n}\}$ is the spatial pattern comprising the ignition points of wildfires registered within a given region W. Characterizing the spatial structure of wildfires and, consequently identifying the areas with higher and lower incidence of wildfires. may be very helpful in the development of fire prevention and fire fighting plans. As in this example, in other scientific areas characterizing the spatial distribution of events is a key factor. In addition, the intensity function needs to be estimated to estimate second-order characteristics such as the K-function and the pair correlation function of inhomogeneous point processes. For these reasons, the estimation of the first-order intensity, $\lambda(x)$ (2.2.1), is a main issue in the analysis of spatial point processes.

To estimate the first-order intensity we can assume a parametric structure and estimate the unknown parameters, using maximum pseudolikelihood or the method of moments (Diggle, 2003; Illian et al., 2008; Waagepetersen, 2007). However, unreliable estimates can be obtained if the assumed parametric model deviates from the true intensity function. Alternatively, the first-order intensity can be estimated nonparametrically. Following the ideas of density estimation for random variables Diggle (1985) introduced the kernel intensity estimator for point processes in $\mathbb{R}$, which can be easily extended to the spatial framework

$$\hat{\lambda}_h(x) = \frac{1}{p_h(x)} \sum_{i=1}^{N} k_h\left(x - \mathbf{x_i}\right) = \frac{1}{h^2 p_h(x)} \sum_{i=1}^{N} k\left(\left(x - \mathbf{x_i}\right)/h\right) \tag{3.1.1}$$

where the kernel function, $k(\cdot)$, is a radially symmetric bivariate probability density function (pdf), $h > 0$ is the smoothing parameter or bandwidth, $k_h(\cdot)$ is the smoothed kernel and $p_h(x) = \int_W h^{-2} k((x-y)/h)dy$ is the edge-correction term. Spatial point patterns are observed on a bounded region, W, that defines a bounded support for the first-order intensity, introducing a discontinuity in the intensity function, which is assumed to be 0 outside W, and the well-known boundary or edge effect. Thus we need to introduce the edge-correction term to overcome this problem and guarantee that $\hat{\lambda}_h(x)$ is asymptotically unbiased.

Although the bandwidth is a key factor in kernel estimation, little attention has been paid to bandwidth selection in spatial point processes. For stationary Cox processes in $\mathbb{R}$, Diggle and Marron (1988) showed that the bandwidth minimizing the mean squared error (MSE) of the kernel intensity estimator equals that provided by least-squares cross-validation for density estimation (Silverman, 1986), and Berman and Diggle (1989) proposed a data-driven procedure to select this bandwidth. This equivalence has not been proved for inhomogeneous point processes. Brooks and Marron (1991) showed that the bandwidth selected by least-squares cross-validation is asymptotically optimal for inhomogeneous point processes in $\mathbb{R}$. Some authors have used smooth bootstrap procedures to select the optimal bandwidth for the kernel density estimator (Taylor, 1989; Cao, 1993; Cao et al., 1994) and the kernel hazard rate estimator (González-Manteiga et al., 1996) and suggested that this approach can perform better than cross-validation. In the point process framework, Loh and Jang (2010) used nonparametric bootstrap to select the optimal bandwidth for the kernel estimator of the pair correlation function, and Cowling et al. (1996) developed a smooth bootstrap technique to create confidence bands for the kernel intensity estimator of inhomogeneous point processes in $\mathbb{R}$. However, to our knowledge bootstrap bandwidth selection procedures have not been proposed for kernel intensity estimation.

The main drawback of the nonparametric intensity estimator is its lack of consistency. As pointed out by Guan (2008b), if we consider a kernel $k(\cdot)$ with finite support around the origin only local information around each point is used to estimate the intensity. If the true intensity is continuous, local smoothing will provide an asymptotically unbiased estimator. But, as the number of events in any region is of order 1, the variance of the estimate does not tend to 0 (see details in Appendix A.1). To overcome this

problem Cucala (2006) defined the *density of event locations* $\lambda_0(x) = \lambda(x)/m$, where $m = \int_W \lambda(x)dx$ is the expected number of events of $\mathbf{X}$ observed on W, and showed the consistency of its kernel estimator. In addition, Guan (2008b) developed a consistent nonparametric intensity estimator assuming that the first-order intensity of a spatial point process is a continuous function of some observed covariates.

Kernel intensity estimation for spatial point processes has been addressed assuming a scalar bandwidth parameter (Diggle, 2003; Cucala, 2008; Guan, 2008b; Illian et al., 2008; Comas et al., 2009), which can be quite restrictive specially for anisotropic and highly inhomogeneous point processes. Following the philosophy of bivariate kernel intensity estimation we can consider bandwidth matrices (Wand, 1992; Wand and Jones, 1994; Duong and Hazelton, 2003) and define the kernel intensity estimator as

$$\hat{\lambda}_H(x) = \frac{1}{p_H(x)} \sum_{i=1}^{N} k_H\left(x - \mathbf{x_i}\right) = p_h(x)^{-1}|H|^{-1/2} \sum_{i=1}^{N} k\left(H^{-1/2}\left(x - \mathbf{x_i}\right)\right) \qquad (3.1.2)$$

where the bandwidth matrix, H, is symmetric and positive-definite and $|H|$ is the determinant of H. Note that expression (3.1.1) corresponds to $H = h^2 I_2$, where I_2 is the 2×2 identity matrix.

The main goal of this chapter is to develop an effective method of bootstrap bandwidth selection for the consistent kernel intensity estimator of spatial point processes. Section 3.2 introduces the density of event locations defined by Cucala (2006) and analyzes the asymptotic behavior of the consistent kernel intensity estimator with full bandwidth matrices. In Section 3.3 we extend to $\mathbb{R}^2$ the smooth bootstrap procedure proposed by Cowling et al. (1996) and show that the bootstrap MISE is a consistent estimator of the MISE of the consistent kernel intensity estimator. This result validates the resampling method proposed. In Section 3.4 we propose a bandwidth selection procedure based on minimizing the bootstrap MISE and analyze its performance by comparison with other methods currently applied. In Section 3.6 we apply the proposed procedure to estimate the first-order intensity of the wildfires registered in Galicia (NW Spain) during 2006. The chapter ends with some conclusions in Section 3.7. Advancements made in this chapter have been published in Fuentes-Santos et al. (2015).

3.2 The consistent kernel intensity estimator

Let $(\mathbf{x_1}, \ldots, \mathbf{x_N})$ be a realization of the inhomogeneous spatial point process X observed on W. Hereafter we assume that the point process is Poisson, i.e. that the events are independent. The Poisson assumption is needed to develop the asymptotic theory and guarantee the consistency of the kernel estimator of the density of event locations. However both the kernel estimator and the bandwidth selector proposed in this chapter can be applied to non-Poisson point processes. In the simulation study (Section 3.5) we analyze their performance for both Poisson and non-Poisson point processes. Notice that we can not distinguish between hetereogeneity and interaction in an observed point pattern unless we have some additional information, such as covariates or a parametric model (Diggle, 2003). Thus, the common practice in the analysis of spatial point patterns is assuming that the point process is Poisson, estimating the first-order intensity function, and then estimating the second-order properties to test the Poisson assumption. Therefore in practice the Poisson assumption is lees restrictive than it could be thought in a first sight.

Given that the number of events of an inhomogeneous Poisson point process, N, has distribution $Poisson(\int_W \lambda(x)dx) = Poisson(m)$, we can establish the following relationship between a bivariate density and the first-order intensity function of a spatial point process

$$f(x) = \frac{\lambda(x)}{\int_W \lambda(x)dx} = \frac{\lambda(x)}{m}$$

Considering this relationship, Cucala (2006) defined the *density of event locations* as $\lambda_0(x) = \lambda(x)/m$. Let H be a symmetric and positive-definite matrix, the kernel estimator of $\lambda_0(\cdot)$ is given by

$$\begin{aligned} \hat{\lambda}_{0,H}(x) &= \frac{\hat{\lambda}_H(x)}{N} I[N \neq 0] \\ &= (p_H(x)N)^{-1} |H|^{-1/2} \sum_{i=1}^{N} k\left(H^{-1/2}(x - \mathbf{x_i})\right) I[N \neq 0] \end{aligned} \tag{3.2.1}$$

A common global error criterion to measure the performance of a kernel estimator is the Mean Integrated Squared Error (MISE). This is the mean integrated squared distance

between an estimate and the target function. In our case

$$MISE(H) = E\left[\int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx\right]$$
$$= \int_W B\left(x, H\right)^2 dx + \int_W Var\left(x, H\right) dx \tag{3.2.2}$$

where E denotes expectation over the randomness in both the location and the number of events, and $B(x, H)$, $Var(x, H)$ are the bias and variance of $\hat{\lambda}_{0,H}(x)$, respectively.

To obtain the asymptotic expression of $MISE(H)$, we need to assume an asymptotic framework. For kernel density estimation in $\mathbb{R}^d$ the typical asymptotics assume that the deterministic sample size, n, tends to infinity and all entries of the bandwidth matrix H tend to 0 such that $n|H|^{1/2} \to \infty$. In the analysis of spatial point patterns two asymptotic frameworks can be assumed: the *increasing-domain* and the *infill or increasing-intensity asymptotics*. The former (Cressie, 1993; Guan, 2008b) dictates that the expected number of events tends to infinity with the size of the observation region. This framework, which increases the number of events keeping the distance between them, can be assumed to analyze the second-order structure of spatial point patterns (Guan, 2007) and to estimate the first-order intensity if we have additional information such as a parametric model (Guan, 2008a) or covariates (Guan, 2008b). However, under this assumption we only add information in the boundary, instead of elsewhere, and when all entries in H tend to zero, the estimated intensity at each point depends on an expected number of events tending to 0. To overcome this problem, Diggle and Marron (1988) adopted the infill asymptotic framework which dictates that the expected number of events tends to infinity, i.e. $\int_{\mathbb{R}^2} \lambda(x)dx \to \infty$, or $m = \int_W \lambda(x) \to \infty$ for a bounded observation domain. Thus, at each location the first-order intensity is estimated using a number of events that tends to ∞. Therefore, the asymptotic theory of this chapter shall be developed under infill asymptotics. This framework has also been assumed by Brooks and Marron (1991) and Cowling et al. (1996) for kernel intensity estimation in $\mathbb{R}$, and by Cucala (2006), which showed the consistency of (3.2.1) with a scalar bandwidth and provided an asymptotic expression for its mean integrated squared error on $\mathbb{R}^2$. Increasing-domain asymptotics can be useful in epidemiology, when both the number of events and the study area increases as a disease spreads. However, the infill asymptotic framework is more appropriate when, as in the analysis of wildfires point patterns, the region is fixed in advance and the number of events increases over time.

Diggle and Marron (1988) pointed out that the edge-correction term in (3.2.1) can introduce a bias towards undersmoothing in bandwidth selection. To overcome this difficulty,

these authors proposed to estimate and minimize the mean squared error the interior of the observation domain in order to obtain the optimal bandwidth. However, the application of this idea implies the selection of a new parameter to define the buffer zone where the error measurement is computed. It should be noted that the expressions of the asymptotic MISE in the interior of a bounded region and in the unbounded domain, $\mathbb{R}^2$, are equal. Therefore, in order to reduce the bias towards undersmoothing and avoid the selection of a new parameter, bandwidth selection shall be based on the expression of the AMISE on $\mathbb{R}^2$. For this reason, throughout this chapter we consider $W = \mathbb{R}^2$. In Appendix A.3 we show the consistency of $\hat{\lambda}_{0,H}(\cdot)$ in a bounded domain, $W \subset \mathbb{R}^2$, and analyze its rate of convergence in the interior and boundary of W.

Before analyzing the performance of (3.2.1) some notation needs to be introduced. For any matrix A, $vechA$ is a column vector containing the lower triangular half of A, dgA denotes the diagonal matrix formed by replacing all non-diagonal entries of A by 0, and $tr(A)$ denotes the sum of its diagonal entries. For any square integrable function g, $R(g) = \int_{\mathbb{R}^2} g(x)^2 dx$. Let $r = (r_1, r_2)^T$, where r_1 and r_2 are non-negative integers, and $|r| = r_1 + r_2$ then the rth partial derivative of λ_0 can be written as

$$\lambda_0^{(r)}(x) = \frac{\partial^{|r|}\lambda_0(x)}{\partial x_1^{r_1}\partial x_2^{r_2}}$$

Denote by $D^r\lambda_0$ the rth derivative of λ_0. Using this notation $D\lambda_0$ and $D^2\lambda_0$ are the gradient vector and Hessian matrix of λ_0. In addition, Ψ_4 is the 3×3 matrix given by

$$\Psi_4 = \int_{\mathbb{R}^2} vech\left[2D^2\lambda_0(x) - dgD^2\lambda_0(x)\right] vech^T\left[2D^2\lambda_0(x) - dgD^2\lambda_0(x)\right] dx$$

where, for any $j = |r|$, each entry in Ψ_j is the integrated density derivative functional

$$\psi_r = \int_{\mathbb{R}^2} \lambda_0^{(r)}(x)\lambda_0(x)dx = E\left[\lambda_0^{(r)}(x)\right]$$

here E denotes expectation over the bivariate density $\lambda_0(x)$. Using integration by parts we can see that $\psi_r \neq 0$ only if $|r|$ is even. For $j = 4$

$$\Psi_4 = \begin{pmatrix} \psi_{40} & 2\psi_{31} & \psi_{22} \\ 2\psi_{31} & 4\psi_{22} & 2\psi_{13} \\ \psi_{13} & 2\psi_{22} & \psi_{04} \end{pmatrix}$$

3.2.1 Regularity conditions and preliminary results

Our results rely on the following assumptions on the bandwidth matrix, the first-order intensity and the kernel function:

I.1 The bandwidth matrix H is symmetric and positive-definite and such that all entries of $H \to 0$ and $m^{-1}|H|^{-1/2} \to 0$, as $m \to \infty$.

I.2 $\lambda(\cdot)$ has partial derivatives up to order 4, all its 2th and 4th order partial derivatives are bounded, continuous and square integrable.

I.3 $k(\cdot)$ is a continuous, symmetric, square integrable density function such that $\int_{\mathbb{R}^2} uu^T k(u)du = \mu_2(k)I_2$, with $\mu_2(k) < \infty$.

Furthermore, given a realization of an inhomogeneous spatial Poisson point process, $\mathbf{X}$, for any measurable function $g(\cdot)$ we have

$$E\left[\frac{1}{N}\sum_{i=1}^{N} g(X_i)I[N \neq 0]\right] = \left(1 - e^{-m}\right)\int_W g(x)\lambda_0(x)dx \tag{3.2.3}$$

and

$$\begin{aligned} Var\left[\frac{1}{N}\sum_{i=1}^{N} g(X_i)I[N \neq 0]\right] &= A(m)\int_W g(x)^2\lambda_0(x)dx \\ &\quad -(A(m) - e^{-m} + e^{-2m})\left(\int_W g(x)\lambda_0(x)dx\right)^2 \end{aligned} \tag{3.2.4}$$

where $A(m) = E\left[\frac{1}{N}I[N > 0]\right] = e^{-m}\sum_{k=1}^{\infty}\frac{m^k}{kk!} < e^{-m}\sum_{k=0}^{\infty}\frac{2m^k}{(k+1)!} = 2/m \to 0$ when $m \to \infty$ (see details in Appendix A.2).

3.2.2 Error measurement and optimal bandwidth for $\hat{\lambda}_{\mathbf{0,H}}(\mathbf{x})$

Under conditions I.2-I.3 and considering $g(X_i) = k_H(x - \mathbf{x}_i)$ in (3.2.3) and (3.2.4), we obtain the following expressions for the bias and variance of $\hat{\lambda}_{0,H}(x)$

$$B(x,H) = -e^{-m}\lambda_0(x) + (1 - e^{-m})\left[\frac{1}{2}\mu_2(k)tr\left(HD^2\lambda_0(x)\right) + o(tr(H))\right] \tag{3.2.5}$$

$$Var\,(x,H) = A(m)|H|^{-1/2}\lambda_0(x)R(k) + o(A(m)|H|^{-1/2}) \tag{3.2.6}$$

Expressions (3.2.5) and (3.2.6) yield

$$MISE(H) = AMISE(H) + o\left(A(m)|H|^{-1/2} + tr^2(H)\right)$$

where

$$AMISE(H) = \frac{1}{4}\mu_2(k)^2 \int_{\mathbb{R}^2} tr^2\left(HD^2\lambda_0(x)\right)^2 dx + A(m)|H|^{-1/2}R(k) \tag{3.2.7}$$

By I.2 all entries in the Hessian matrix of $\lambda_0(x)$, $D^2\lambda_0(x)$, are square integrable, and by I.3 $R(k) < \infty$ and $\mu_2(k) < \infty$. Thus $AMISE(H)$ tends to 0 when $m \to \infty$ provided that by I.1 all entries of $H \to 0$ and $m^{-1}|H|^{-1/2} \to 0$ (see details and expressions for $W \subset \mathbb{R}^2$ in Appendix A.3).

An alternative form of the AMISE is given by

$$AMISE(H) = \frac{1}{4}\mu_2(k)^2 (vech^T H)\Psi_4(vechH) + A(m)|H|^{-1/2}R(k) \tag{3.2.8}$$

As stated by Marron and Wand (1992) for kernel density estimator, the remainder term in the MISE can be considerably large for point processes with low mean intensity. However, we have observed that even when the AMISE is a poor estimator of the MISE, the discrepancy between H_{AMISE} and H_{MISE} is small (see Appendix A.4). This supports the use of the AMISE as criterion for bandwidth selection.

The minimization of $AMISE(H)$ can only be performed numerically. To do so we can apply a Newton-Raphson algorithm, as proposed by Wand (1992) for multivariate kernel density estimators. If we restrict H to the family of positive-definite diagonal matrices, i.e. $H = diag(h_1^2, h_2^2)$, we obtain a simplified expression for (3.2.8). The AMISE of $\hat{\lambda}_{0,h}(x)$ for $h = (h_1, h_2)^T$ is

$$AMISE(H) = \frac{1}{4}\mu_2(k)^2\left(h_1^4\psi_{40} + 2h_1^2h_2^2\psi_{22} + h_2^4\psi_{04}\right) + \frac{A(m)}{h_1h_2}R(k) \tag{3.2.9}$$

and the bandwidth vector minimizing (3.2.9) is $h_{AMISE} = (h_{1,AMISE}, h_{2,AMISE})^T$ where

$$\begin{aligned} h_{1,AMISE} &= \left[\frac{\psi_{04}^{3/4}A(m)R(k)}{\mu_2(k)^2\psi_{40}^{3/4}\left(\psi_{40}^{1/2}\psi_{04}^{1/2} + \psi_{22}\right)}\right]^{1/6} \\ h_{2,AMISE} &= h_{1,AMISE}\left(\psi_{40}/\psi_{04}\right)^{1/4} \end{aligned} \tag{3.2.10}$$

Finally, for $H = h^2 I_2$ we have

$$h_{AMISE} = \left[\frac{2A(m)R(k)}{\mu_2(k)^2\left(\psi_{40} + 2\psi_{22} + \psi_{04}\right)}\right]^{1/6} \tag{3.2.11}$$

(see details in Appendix A.5).

The AMISE of $\hat{\lambda}_{0,H}(x)$, (3.2.8), depends on m and Ψ_4 that, as well as the first-order intensity, are unknown. Thus, in order to use the MISE as error criterion to select the optimal bandwidth, we need to obtain an estimator of $AMISE\,(H)$.

3.3 A bootstrap kernel intensity estimator

3.3.1 Resampling procedure

Nonparametric bootstrap procedures, which generate replicates of a dataset, allow performing inference and developing goodness-of-fit tests in different areas of Statistics. Bootstrap techniques have been used with great success with independent data, and have been adapted to the case of spatial or temporal data involving dependent observations. For example Lahiri et al. (1999) applied bootstrap to make inference on the spatial cumulative distribution function, and Guan and Loh (2007) applied it to fit models to stationary point patterns. Marked-points bootstrap was used by Loh and Stein (2004) and Loh (2010) to test for the goodness-of-fit of the K-function of homogeneous and inhomogeneous point patterns, and by Loh and Jang (2010) to select the bandwidth for the nonparametric estimator of the pair correlation function. Here we extend to $\mathbb{R}^2$ the smooth bootstrap procedure introduced by Cowling et al. (1996) for inhomogeneous Poisson point processes in $\mathbb{R}$ in order to estimate the MISE of the kernel estimator of the density of event locations given by (3.2.1).

Let $(\mathbf{x_1}, \ldots, \mathbf{x_N})$ be a realization of the inhomogeneous spatial Poisson point process with first-order intensity $\lambda(x)$ introduced above, and let $\hat{\lambda}_G(x)$ be a kernel intensity estimator with pilot bandwidth matrix G. The smooth bootstrap resamples can be generated by two equivalent procedures (Devroye and Györfi, 1985):

Method 1

1. Conditional on $(\mathbf{x_1}, \ldots, \mathbf{x_N})$, let $N^* \sim Poisson\left(\int_W \hat{\lambda}_G(x)dx\right)$.

2. Draw $(\mathbf{x_1^*}, ..., \mathbf{x_{N^*}^*})$ by sampling randomly with replacement N^* times from the distribution with density $\hat{\lambda}_G(x)/\int_W \hat{\lambda}_G(x)dx$.

Method 2

1. Conditional on $(\mathbf{x_1}, \ldots, \mathbf{x_N})$, let $N^* \sim Poisson\left(\int_W \hat{\lambda}_G(x)dx\right)$.
2. The resampled point pattern is defined as: $\mathbf{x_i^*} = \mathbf{y_i^*} + z_i^*, i = 1, \ldots, N^*$, where $\mathbf{y_i^*}$ is drawn by sampling randomly with replacement from $(\mathbf{x_1}, \ldots, \mathbf{x_N})$, the z_i^*´s are independent and identically distributed with a bivariate density $k(\cdot)$ rescaled by a smoothing matrix G, which determines the disturbance applied to the events.

As pointed out by Cowling et al. (1996), the kernel and bandwidth used in the smoothing step do not need to be the same as those used to estimate $\lambda(x)$. In this work Gaussian kernels are used for both steps, i.e. $Z^* \sim N_2(0, G)$.

3.3.2 Bootstrap kernel intensity estimator

Let $(\mathbf{x_1^*}, \cdots, \mathbf{x_{N^*}^*})$ be a bootstrap resample of the spatial point pattern $\{\mathbf{x_i}\}_{i=1}^N$ observed on $W \subset \mathbb{R}^2$. The bootstrap version of the kernel estimator of $\lambda_0(x)$ given in (3.2.1) with bandwidth matrix H is

$$\begin{aligned}\hat{\lambda}_{0,H}^*(x) &= \frac{1}{N^*}\hat{\lambda}_H^*(x) I\left[N^* > 0\right] \\ &= \left(p_H\left(x\right) N^*\right)^{-1} |H|^{-1/2} \sum_{i=1}^{N^*} k\left(H^{-1/2}(x - \mathbf{x_i^*})\right) I\left[N^* > 0\right]\end{aligned} \tag{3.3.1}$$

As well as for $\hat{\lambda}_{0,H}(x)$, the performance of $\hat{\lambda}_{0,H}^*(x)$ shall be measured in terms of its MISE, given by

$$\begin{aligned}MISE^*(H) &= E^*\left[\int_W \left(\hat{\lambda}_{0,H}^*(x) - \hat{\lambda}_{0,G}(x)\right)^2 dx\right] \\ &= \int_W B^*(x,H)^2 dx + \int_W Var^*(x,H)dx\end{aligned} \tag{3.3.2}$$

where $\hat{\lambda}_{0,G}(x)$ stands for the kernel estimator of $\lambda_0(x)$ used to generate the bootstrap pattern. The following theorem provides the asymptotic expression for $MISE^*(H)$ in the unbounded domain and establishes the conditions that guarantee its consistency as an estimator of $MISE(H)$ (see proof in Appendix A.6).

Theorem 3.1. *Let $\hat{\lambda}_{0,H}(x)$ be the bootstrap kernel estimator of the density of event locations for an inhomogeneous spatial Poisson point process X. Under conditions I.1-I.3:*

$$MISE^*(H) = AMISE^*(H) + o_p\left(tr^2(H) + A(\hat{m})|H|^{-1}\right) \tag{3.3.3}$$

where

$$AMISE^*(H) = \frac{1}{4}\mu_2(k)^2(vech^T H)\hat{\Psi}_{4,G}(vechH) + A(\hat{m})|H|^{-1/2}R(K) \tag{3.3.4}$$

tends to 0 when $m \to \infty$. Thus $\hat{\lambda}^*_{0,H}(x)$ is a consistent estimator of $\hat{\lambda}_{0,G}(x)$. Furthermore, $AMISE^*(H)$ is a consistent estimator of $AMISE(H)$ if $\hat{\Psi}_{4,G}$ is a consistent estimator of Ψ_4.

Comparison of expressions (3.2.8) and (3.3.4) shows that the dominant terms of the integrated square bias and the integrated variance in the bootstrap versions are estimators of the corresponding terms in $AMISE(H)$. Note that $A(m)$ can be estimated by $\hat{A}(m) = 1/n$. To estimate Ψ_4 we should take into account that $\psi_r = E\left[\lambda_0^{(r)}(x)\right]$. Therefore, if we consider the kernel estimator of $\lambda_0^{(r)}(x)$

$$\hat{\lambda}^{(r)}_{0,G}(x) = N^{-1}\sum_{i=1}^{N} k_G^{(r)}\,(x - \mathbf{x_i})I[N \neq 0] \tag{3.3.5}$$

and use property (3.2.3) for $g(\mathbf{x_i}) = k_G^{(r)}(x - \mathbf{x_i})$ we have

$$\hat{\psi}_{r,G} = E\left[\hat{\lambda}^{(r)}_{0,G}(x)\right] = (1 - e^{-m})\int_{\mathbb{R}^2} k_G^{(r)}(x - y)\lambda_0(y)dy \tag{3.3.6}$$

where E denotes expectation over the randomness in the point pattern. Conditional on $N = n$ the observed pattern, $(\mathbf{x_1}, ..., \mathbf{x_n})$, can be seen as a random sample of size n of the bivariate random variable with density $f(x) = \lambda_0(x)$. Then, the integral in the right hand side of expression (3.3.6) is the mean of the kernel estimator of $\psi_r(f)$ for a bivariate density function f, which is given by $\hat{\psi}_{r,G}(f) = n^{-2}\sum_{i=1}^{n}\sum_{j=1}^{n} k_G(\mathbf{x_i} - \mathbf{x_j})$. Thus

$$E\left[\hat{\psi}_{r,G}\right] = (1 - e^{-m})E\left[\hat{\psi}_{r,G}(f)\right]$$

and

$$Var\left[\hat{\psi}_{r,G}\right] = (1 - e^{-m})^2 Var\left[\hat{\psi}_{r,G}(f)\right]$$

Therefore, we can derive the sample properties of $\hat{\psi}_{4,G}$ from those obtained by Duong and Hazelton (2003) for the bivariate kernel density estimator. If we add some extra regularity conditions to $\lambda(\cdot)$ and $k(\cdot)$, replacing conditions I.2, I.3 by

I.2b $\lambda(x)$ has partial derivatives up to order $j+2$, all its jth and $(j+2)$th order partial derivatives are bounded, continuous and square integrable,

I.3b $k(\cdot)$ is a continuous, symmetric, square integrable density function such that $\int_{\mathbb{R}^2} uu^T k(u)du = \mu_2(k)I_2$, with $\mu_2(k) < \infty$, and all its jth order partial derivatives are square integrable,

we can obtain the sample properties of $\hat{\psi}_{r,G}$ for any even $j = |r|$. As the expressions for a general G are intractable, we consider a scalar pilot bandwidth, i.e. $G = g^2 I_2$, and obtain the following expression for the bias of $\hat{\psi}_{r,g}$

$$B\left[\hat{\psi}_{r,g}\right] = -e^{-m}\psi_r + (1-e^{-m})n^{-1}g^{-|r|-2}k^{(r)}(0) + \frac{(1-e^{-m})}{2}g^2\mu_2(k)\sum_{j=1}^{2}\psi_{r+2e_i} + o((1-e^{-m})g^2)$$

where e_i is the ith elementary vector. The variance of $\hat{\psi}_{r,g}$ is given by

$$Var\left[\hat{\psi}_{r,g}\right] = 2n^{-2}g^{-2|r|-2}(1-e^{-m})^2\psi_0 R(k^{(r)}) + o\left((1-e^{-m})^2(n^{-2}g^{-2|r|-2} + n^{-1})\right)$$

provided that $k^{(r)}$ is square integrable, $g \to 0$ and $m^{-2}g^{-2|r|-2} \to 0$ as $m \to \infty$. When $m \to \infty$ we obtain the asymptotic mean squared error of $\hat{\psi}_{r,g}$

$$AMSE\left(\hat{\psi}_{r,g}\right) = 2n^{-2}g^{-2|r|-2}\psi_0 R(k^{(r)}) + \left[n^{-1}g^{-|r|-2}k^{(r)}(0) + \frac{g^2}{2}\mu_2(k)\sum_{i=1}^{2}\psi_{r+2e_i}\right]^2 \tag{3.3.7}$$

which is equal to $AMSE(\hat{\psi}_{r,g}(f))$. Although hereafter our bandwidth selection criterion equals the plug-in procedure for bivariate density estimation, we reproduce the expressions in Duong and Hazelton (2003) as they are useful to a complete understanding of this work. We use the same bandwidth, g, to estimate all terms in Ψ_4 instead of selecting an optimal bandwidth for each term (Wand and Jones, 1994), which can lead to non positive-definite bandwidth matrices. The optimal bandwidth is such that minimizes

the sum of the AMSEs (SAMSE) for $\hat{\psi}_{r,g}$

$$SAMSE(\hat{\Psi}_{j,g}) = \sum_{r,|r|=j} AMSE_0(\psi_{r,g}) \tag{3.3.8}$$
$$= 2n^{-2}g^{-2j-2}A_1 + 2n^{-2}g^{-2j-4}A_2 + n^{-1}g^{-j}A_3 + \frac{1}{4}g^4A_4$$

where

$$A_1 = \sum_{r,|r|=j} R(k^{(r)})$$

$$A_2 = \sum_{r,|r|=j} k^{(r)}(0)^2$$

$$A_3 = \mu_2(k) \sum_{r,|r|=j} k^{(r)}(0)^2 \left(\sum_{i=1}^{2} \psi_{r+2e_i} \right)$$

and

$$A_4 = \mu_2(k)^2 \sum_{r,|r|=j} \left(\sum_{i=1}^{2} \psi_{r+2e_i} \right)^2$$

A_1, A_2 and A_4 are positive. In addition, if k is a Gaussian kernel $A_3 < 0$ given that when all elements of r are even, $k^{(r)}(0)$ and ψ_{r+2e_i} are of opposite signs, and when at least one element is odd, then $k^{(r)}(0) = 0$. Finally, as the first term in (3.3.8) is dominated by the second, we can remove the former, which is the asymptotic variance, and obtain

$$SAMSE(\hat{\Psi}_{j,g}) = 2n^{-2}g^{-2j-4}A_2 + n^{-1}g^{-j}A_3 + \frac{1}{4}g^4A_4 \tag{3.3.9}$$

Minimizing (3.3.9) we obtain the rth order SAMSE-optimal bandwidth

$$g_{SAMSE} = \left[\frac{(4j+8)A_2}{(-jA_3 + \sqrt{j^2A_3^2 + (8j+16)A_2A_4})n} \right]^{1/(j+4)} \tag{3.3.10}$$

Considering the regularity assumptions I.2b-I.3b for $j = 4$, we have that the SAMSE of $\hat{\Psi}_{4,g}$ has order $O(n^{-1/2}) = O(m^{-1/2})$ and tends to 0 as $m \to \infty$. Thus $AMISE^*(H)$ is a consistent estimator of $AMISE(H)$. This result provides a formal proof for the validity of the smooth bootstrap procedure. According to (3.3.10) the pilot bandwidth has order $O(m^{-1/8})$, then the bootstrap resampling should be conducted with a bandwidth asymptotically larger than the optimal bandwidth for $\hat{\lambda}_{0,H}(x)$, which is $O(m^{-1/6})$. Finally, note that $AMISE^*(H)$ depends on the observed pattern but not on the bootstrap resamples, i.e. we do not need to conduct the bootstrap procedure to estimate the MISE of $\hat{\lambda}_{0,H}(x)$.

3.4 Bandwidth selection

A key factor affecting the accuracy of the kernel intensity estimator of a spatial point pattern is the choice of the smoothing parameter or bandwidth, H. If the bandwidth is too small, the kernel estimators pays too much attention to any particular event leading to an undersmoothing of the intensity function. In contrast, if the bandwidth is too large, the kernel estimator cannot capture the spatial variability of the data leading to an oversmoothing of the intensity function. In this section we discuss the bandwidth selectors currently available, and propose a plug-in bandwidth selection procedure based on the bootstrap MISE of the kernel estimator of $\lambda_0(x)$.

3.4.1 Bandwidth selection procedures

Diggle´s bandwidth selector

Diggle (1985) proposed a bandwidth selector based on minimizing the mean squared error (MSE) of the kernel intensity estimator for stationary and isotropic Cox process. Let $(\mathbf{x}_1, \ldots, \mathbf{x}_n)$ be a partial realization of a stationary Cox process, $\mathbf{X}$, observed in the region $W \subset \mathbb{R}^2$. If the driving process has intensity $\Lambda(x)$ with mean λ and covariance $\gamma(u)$, then, λ and $\lambda_2(u) = \gamma(u) - \lambda^2$ are respectively the first and second-order intensities of $\mathbf{X}$. Considering the uniform kernel and ignoring edge-effects, the intensity estimator (3.1.1) reduces to

$$\hat{\lambda}_h(x) = N(x, h)/(\pi h^2)$$

where, $N(x, h)$ is the number of events of $\mathbf{X}$ within distance h of x. Conditional on the realization of the rate process $\Lambda(\cdot)$, $N(x, h)$ follows a Poisson distribution with mean and variance $\int \Lambda(x)dx$ and $\int \lambda_2(\|x\|)\, dx = \lambda^2 K(h)$. Thus taking $x = 0$ given that under stationarity $MSE(h)$ does not depend on x, the MSE of $\hat{\lambda}_h(x)$ is

$$MSE(h) = \lambda_2(0) + \lambda\left[1 - 2\lambda K(h)\right]/(\pi h^2) + (\pi h^2)^{-2} \int\int \lambda_2(\|x - y\|)dydx$$

as the first term in the right hand side of this equation does not depend on h. Minimizing $MSE(h)$ is equivalent to minimizing

$$M(h) = \lambda\left[1 - 2\lambda K(h)\right]/(\pi h^2) + (\pi h^2)^{-2} \int\int \lambda_2(\|x - y\|)dydx \tag{3.4.1}$$

The first term in the right hand side of this expression depends on the first-order intensity and Ripley´s K-function, which can be substituted by their empirical estimators

$\hat{\lambda} = n/|W|$ and $\hat{K}(h)$ (2.2.4), whilst the double integral can be reduced to a scalar integral using polar coordinates and estimated by the procedure explained in Berman and Diggle (1989).

This bandwidth selector was implemented in the function *bw.diggle* of the *spatstat* package of R (Baddeley and Turner, 2005; R Core Team, 2014) for the isotropic Gaussian kernel. The smoothing parameter returned by *bw.diggle*, which was obtained through a second order calibration of the Gaussian kernel against the uniform kernel, corresponds to $h/2$, where h is the bandwidth parameter minimizing expression (3.4.1).

Diggle and Marron (1988) proved the equivalence between this criterion and least-squares cross-validation for kernel density estimators of random variables. However, this equivalence has not been tested for inhomogeneous point processes which does not belong to the stationary Cox family. In addition, this bandwidth selection procedure was developed under the stationarity assumption. Thus for inhomogeneous point processes the K-function in $M(h)$ is estimated under a wrong hypothesis. So, we need to check the accuracy of this criterion for inhomogeneous Poisson point processes.

Least-squares cross-validation

The optimal bandwidth can be defined as that which minimizes the integrated square error (ISE) of the kernel intensity estimator

$$\begin{aligned} ISE(h) &= \int_W \Big(\hat{\lambda}_h(x) - \lambda(x)\Big)^2 dx \\ &= \int_W \hat{\lambda}_h^2(x)dx - 2\int_W \hat{\lambda}_h(x)\lambda(x)^2 dx + \int_W \lambda^2(x)dx \end{aligned} \tag{3.4.2}$$

Following the proposal of Rudemo (1982) and Bowman (1984) for kernel density estimates, Brooks and Marron (1991) introduced the least-squares cross-validation (LSCV) score function for kernel intensity estimation

$$LSCV(h) = \int_W \hat{\lambda}_h(x)dx - 2\sum_{\mathbf{x}_i \in \mathbf{X}\cap W} \hat{\lambda}_h^{(-i)}(\mathbf{x}_i) \tag{3.4.3}$$

where $\hat{\lambda}_h(x)$ and $\hat{\lambda}_h^{(-i)}(x)$ are the kernel intensity estimator and its cross-validation version obtained by deleting the i-th event of the point pattern. $LSCV(h)$ is a reasonable unbiased estimate of the terms in expression 3.4.2 that depend on h. Therefore, the

bandwidth that minimizes $LSCV(h)$ should be close to the bandwidth that minimizes $ISE(h)$. For intensity functions with continuous and bounded second order derivatives and kernel functions with bounded support Brooks and Marron (1991) showed that the least-squares cross-validation bandwidth selector is asymptotically optimal almost surely.

Following the proposal of Duong and Hazelton (2005) for the bivariate kernel density estimator, this criterion can be extended to the case of diagonal and full bandwidth matrices.

Pseudolikelihood cross-validation

In application to real data, least squares cross-validation can produce unacceptable small bandwidths, leading to an extreme undersmoothing of the first-order intensity. An alternative bandwidth selector is the extension to the spatial point process context of the maximum likelihood cross-validation method for kernel density estimation (Loader, 1999). This extension provides the pseudolikelihood cross-validation (PLCV) criterion, which defines the optimal bandwidth as the maximizer of

$$PLCV(h) = n^{-1}\sum_{i=1}^{n} \log \hat{\lambda}_h^{-i}(x_i) - \int_W \hat{\lambda}_h(x)dx \tag{3.4.4}$$

where $\hat{\lambda}_h(x)$ and $\hat{\lambda}_h^{(-i)}(x)$ are the kernel intensity estimator and its cross-validation version. This procedure was implemented in the function *bw.ppl* of *spatstat* (Baddeley and Turner, 2005).

Rule of thumbs by Scott

In the *spatstat* package we can also find the function *bw.scott* which uses the rule of thumbs proposed by Scott (1992) for kernel density estimation to obtain a diagonal bandwidth matrix, $H = diag\left(h_1^2, h_2^2\right)$ where

$$h_i = sd(X_i)n^{-1/6},\ i = 1, 2 \tag{3.4.5}$$

and X_i, X_2 are respectively, the unidimensional patterns containing the first and second order coordinates of $\mathbf{X}$. This procedure provides larger smoothing parameters than

Diggle´s criterion.

3.4.2 Plug-in bandwidth selector

Bandwidth selection for the kernel estimator of $\lambda_0(x)$, and therefore for the kernel intensity estimator, can be conducted minimizing the bootstrap AMISE given by (3.3.4), to that end we need a pilot estimate of Ψ_4. Furthermore, for any even j (3.3.9) shows that optimal SAMSE smoothing of Ψ_j requires pilot estimates of order $j+2$. This situation suggests conducting a plug-in procedure that implements the SAMSE approach up to a maximum level j_{max} and estimates the $(j_{max}+2)$th order functionals by the following normal reference rule: $\hat{\psi}_r^{NR} = (-1)^{|r|}\phi_{2S}^{(r)}(0)$, where ϕ_{2S} is the bivariate normal density with mean 0 and covariance matrix $\Sigma = 2S$, and S is the covariance matrix of the event locations.

Therefore, following Duong and Hazelton (2003) the optimal bandwidth matrix can be selected by the following l-stage plug-in algorithm:

1. Set $j_{max} = 2l + 4$. Obtain the normal reference estimators $\hat{\psi}_r^{NR}$ for $|r| = j_{max}$. Plug these estimators into the j_{max}th SAMSE optimal bandwidth.
2. For $j = j_{max} - 2,\ j_{max} - 4, \ldots$
   (a) Calculate kernel estimators of ψ_r of order $j = |r|$ using the plug-in bandwidths $g_{j,SAMSE}$
   (b) Substitute $\hat{\psi}_{r,g_{j,SAMSE}}$ into equation (3.3.10) to give plug-in estimates of $g_{j-2,SAMSE}$
3. Use $g_{4,SAMSE}$ to obtain $\hat{\Psi}_{4,g}$. Plug this estimate into equation (3.3.4).
4. Minimize (3.3.4) to obtain the optimal bandwidth matrix H_{AMISE^*}. To this purpose we can apply the Newton-Raphson numerical algorithm.

For ease of computation, this algorithm uses pilot bandwidths of the form $G = g^2 I_2$. However, as indicated for H, these bandwidths can be very restrictive in particular for anisotropic or highly inhomogeneous point patterns. This difficulty can be overcome applying two transformations to the event locations before applying the plug-in algorithm. Let S be the covariance matrix of event locations, we define the pre-scaled data as $\{X_i^*\}_{i=1}^n = \left\{dgS^{-1/2}X_i\right\}_{i=1}^n$, and the pre-sphered data as $\{X_i^*\}_{i=1}^n = \left\{S^{-1/2}X_i\right\}_{i=1}^n$.

The plug-in bandwidth $H^*_{AMISE^*}$ should be back-transformed to the original scale by $H_{AMISE^*} = dgS^{1/2}H^*_{AMISE^*}dgS^{1/2}$ and $H_{AMISE^*} = S^{1/2}H^*_{AMISE^*}S^{1/2}$ for pre-scaled and pre-sphered data, respectively.

3.5 Simulation study

The performance of the plug-in bandwidth selector proposed above was analyzed through its application to simulated spatial Poisson point processes with different levels of inhomogeneity. We generated $B = 100$ realizations of six spatial point processes with $m = 500$ and $m = 1000$ on the unit square. Figures 3.1 and 3.2 show the first-order intensities and a realization of each simulated process, respectively.

The first spatial point process has intensity

$$\lambda_1(x) = bexp(-3x_1)$$

with $b = 1700$ and $b = 3300$ to obtain $m = \int_W \lambda(x)dx = 500$ or 1000 respectively. To generate the second and third point patterns we considered the intensity function

$$\lambda(x) = b\phi_{(0.3-0.2x_2,\sigma)}(x_1) + 25$$

where ϕ is the univariate normal density with mean $\mu = 0.3 - 0.2x_2$ and standard deviation $\sigma = 0.1$ in $\lambda_2(x)$ and $\sigma = 0.02$ in $\lambda_3(x)$, with $b = 500$ and $b = 1000$ to obtain $m = 500$ and $m = 1000$. This model was used by Barr and Schoenberg (2010) to analyze the performance of the Voronoi estimator of the first-order intensity. We also simulated log-Gaussian Cox processes with intensity function

$$\lambda_4(x) = \exp(6 + 4Y(x))$$

where Y is a realization of a Gaussian Random Field with mean 0 and exponential model for the covariance function $C(t) = \sigma^2 exp(-t/\rho), t > 0, \sigma = \rho = 0.1$ (Loh, 2010). This intensity was multiplied by 2 to obtain $m = 1000$. Finally, we simulated spatial point processes with first-order intensities

$$\lambda_5(x) = b\left(\phi_{x_2,0.1}(x_1) + \phi_{1-x_2,0.1}(x_1) - \frac{1}{5}\phi_{2,(\mu,\Sigma)}(x)\right) + 30$$

and

$$\lambda_6(x) = b\left(\phi_{x_2,0.1}(x_1) + \phi_{1-x_2,0.2}(x_1) - \frac{1}{20}\phi_{2,(\mu,\Sigma)}(x)\right) + 50$$

where $\phi_{2,(\mu,\Sigma)}$ is the bivariate normal density with mean $\mu = (0.5, 0.5)$, and variance matrix $\Sigma = 0.01I_2$, In order to obtain $m = 500$ and 1000 we used $b = 275$ and $b = 600$, respectively.

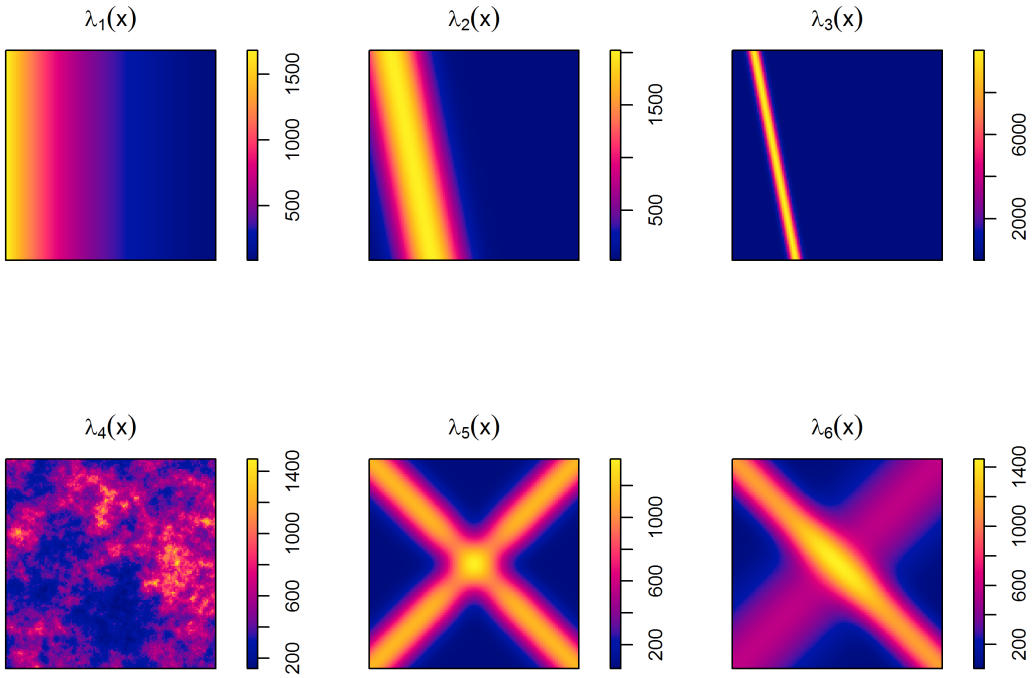


FIGURE 3.1: First-order intensity functions of the inhomogeneous spatial Poisson point processes used in the simulation study ($m = 1000$).

We also tested the performance of our bandwidth selection procedure for Thomas Cluster processes with first-order intensities $\lambda_1(x)$ and $\lambda_2(x)$. In both cases the expected number of parent points was $\kappa = 100$ and $\kappa = 200$ for $m = 500$ and 1000, respectively, and the expected number of offsprings per parent was $\mu = 5$. Offspring were distributed around the parent points according to a Gaussian distribution with $\sigma = 0.1$ for $\lambda_1(x)$ and $\sigma = 0.05$ for $\lambda_2(x)$ (Figure 3.3).

For each simulated point pattern we applied the 1 and 2-stage plug-in bandwidth selector considering a scalar bandwidth parameter (h), diagonal bandwidth matrices (D) and full bandwidth matrices (F). For diagonal bandwidth matrices, the plug-in algorithm was

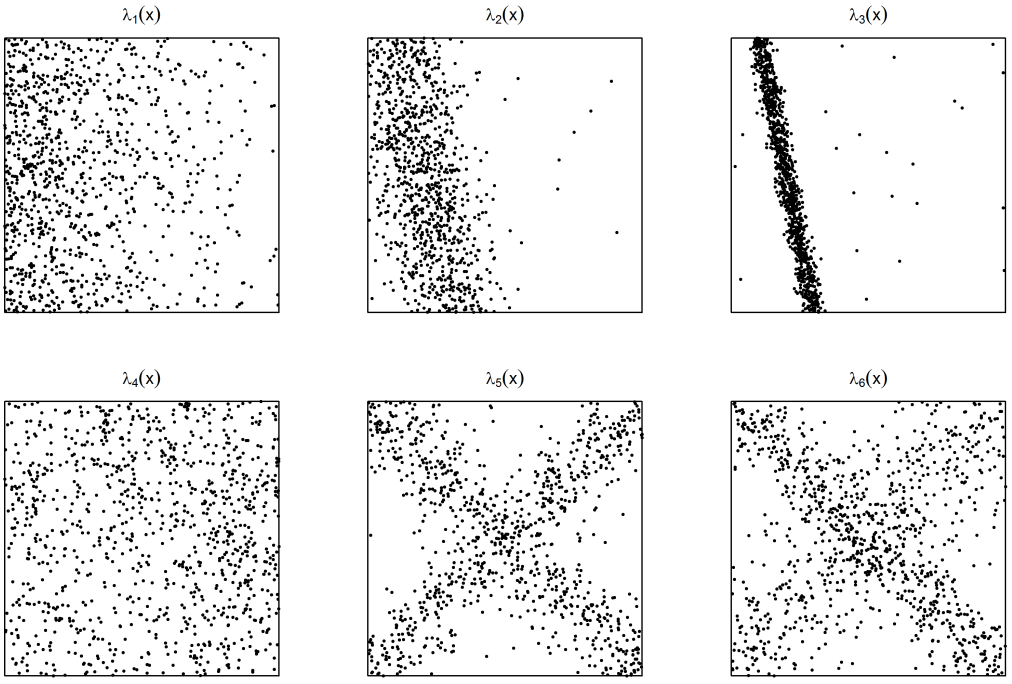


FIGURE 3.2: Realizations of the inhomogeneous spatial Poisson point processes used in the simulation study ($m = 1000$).

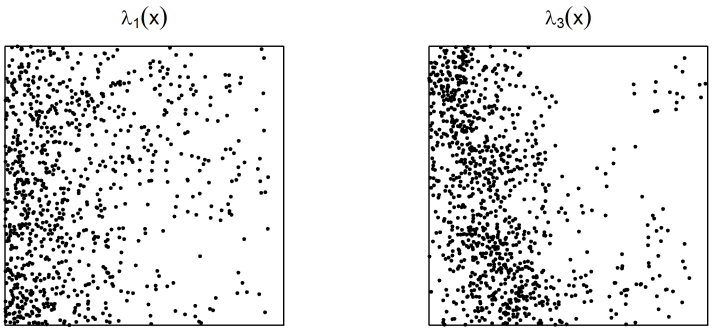


FIGURE 3.3: Realizations of the Thomas Cluster processes with first-order intensities $\lambda_1(x)$ and $\lambda_3(x)$ ($m = 1000$).

applied over raw and pre-scaled (s) event locations, while for full bandwidth matrices event locations were pre-scaled (s) or pre-sphered (sp). In order to obtain the full bandwidth matrix minimizing the AMISE, we applied a Newton-Rapson algorithm, as H is assumed to be symmetric we estimate $vechH$, which contain all the different entries of H. We used $vechH_0$ as a vector of initial parameters in the optimization algorithm, with H_0 the normal scale bandwidth for bivariate density estimators $H_0 = n^{-1/3}S$, where S is the covariance matrix of the event locations, which is positive-definite. All the matrices provided by this procedure were positive-definite as required.

The scalar plug-in bandwidth selector was compared with Diggle´s criterion (Diggle and Marron, 1988), least-squares cross-validation (LSCV) (Brooks and Marron, 1991) and pseudolikelihood cross-validation (PLCV) (Loader, 1999; Baddeley and Turner, 2005), which consider the edge-correction term in the error measurement; the diagonal plug-in bandwidth selector was compared with Scott´s rule of thumbs (Scott, 1992); and the full matrix plug-in was compared with full-matrix least-squares cross-validation (Flscv). Full matrix cross-validation was performed following the proposal of Duong and Hazelton (2005) for the bivariate kernel density estimator, but we do not have prior information about its performance for intensity estimation in spatial point processes. Both Scott´s rule of thumbs and full matrix cross-validation do not consider the edge-correction term.

The efficiency of the bandwidth selectors outlined above was compared by means of the integrated squared error of the kernel estimator of $\lambda_0(x)$ obtained with the respective optimal bandwidths

$$ISE = \int_W \left(\hat{\lambda}_{0,H}(x) - \lambda_0(x)\right)^2 dx$$

The simulation study was conducted with the spatstat (Baddeley and Turner, 2005) and ks (Duong, 2013a) packages of R (R Core Team, 2014).

Figures 3.4 and 3.5 show boxplots for the ISE of the kernel intensity estimators with the bandwidths provided by each selector, we only show the boxplots for $m = 1000$ as the results for $m = 500$ were quite similar. Tables A.2-A.9 in Appendix A.7 show the mean optimal bandwidth provided by each criteria, the corresponding MISE, i.e. the mean of the bandwidths and ISEs obtained for the 100 simulated patterns. We also show the computational cost of each procedure. Scott´s rule of thumbs and Diggle´s criterion registered the lowest computational demand, while full matrix cross-validation was the most intensive. The computational cost of the plug-in bandwidth selectors increased with the number of stages but neither the complexity of the bandwidth matrix nor the transformation of the event locations increased their computational demand.

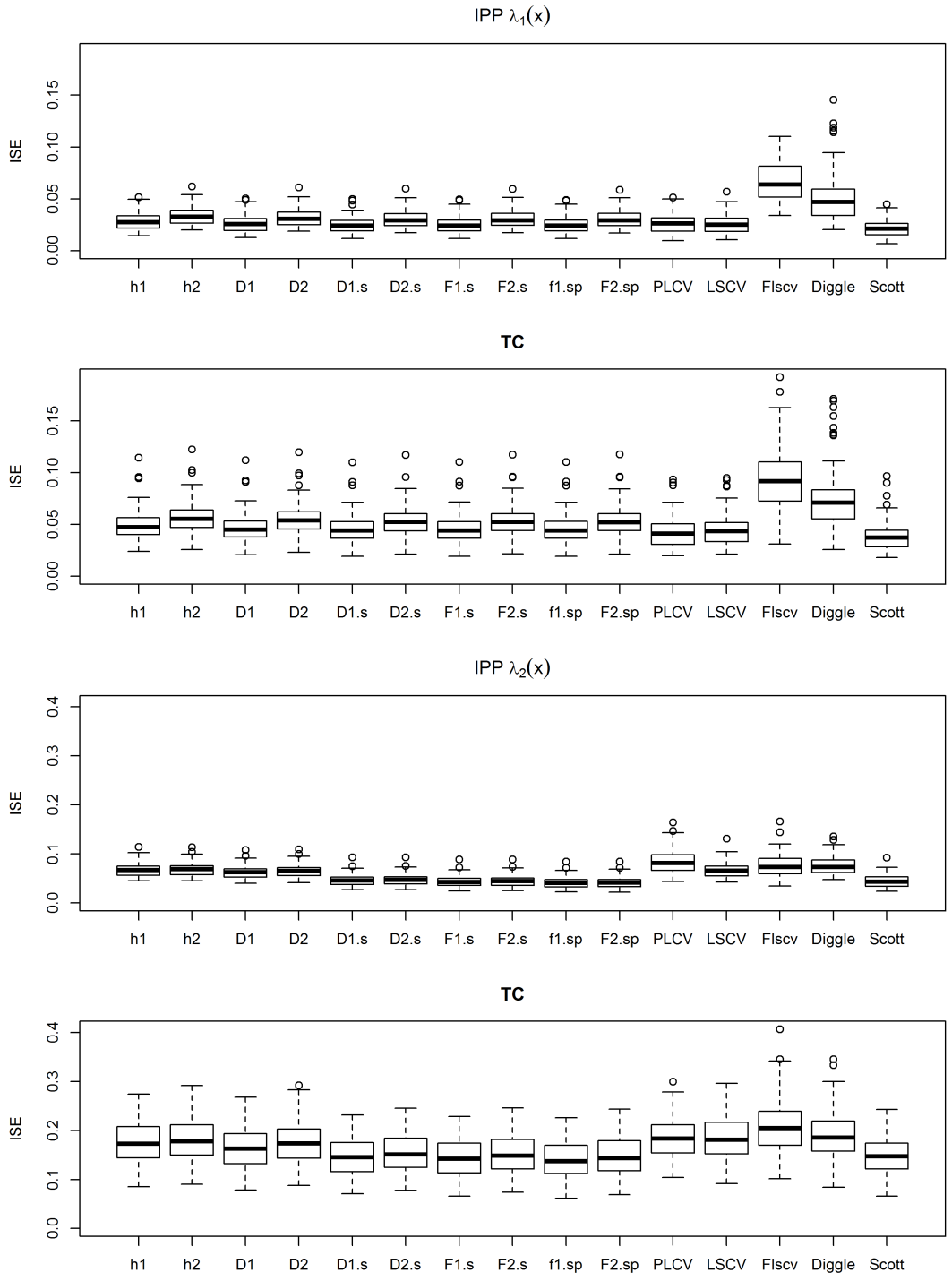


FIGURE 3.4: Comparison of bandwidth selectors for Poisson (IPP) and clustered (TC) point patterns with first-order intensities $\lambda_1(x)$ and $\lambda_2(x)$. Labels "1" and "2" indicate 1 and 2-stage plug-in bandwidth selection; h: scalar bandwidth; D: Diagonal matrix; F: full matrix. "s" indicates pre-scaled and "sp" pre-sphered data; PLCV: pseudo-likelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s bandwidth selector; Scott: Scott´s rule of thumbs.

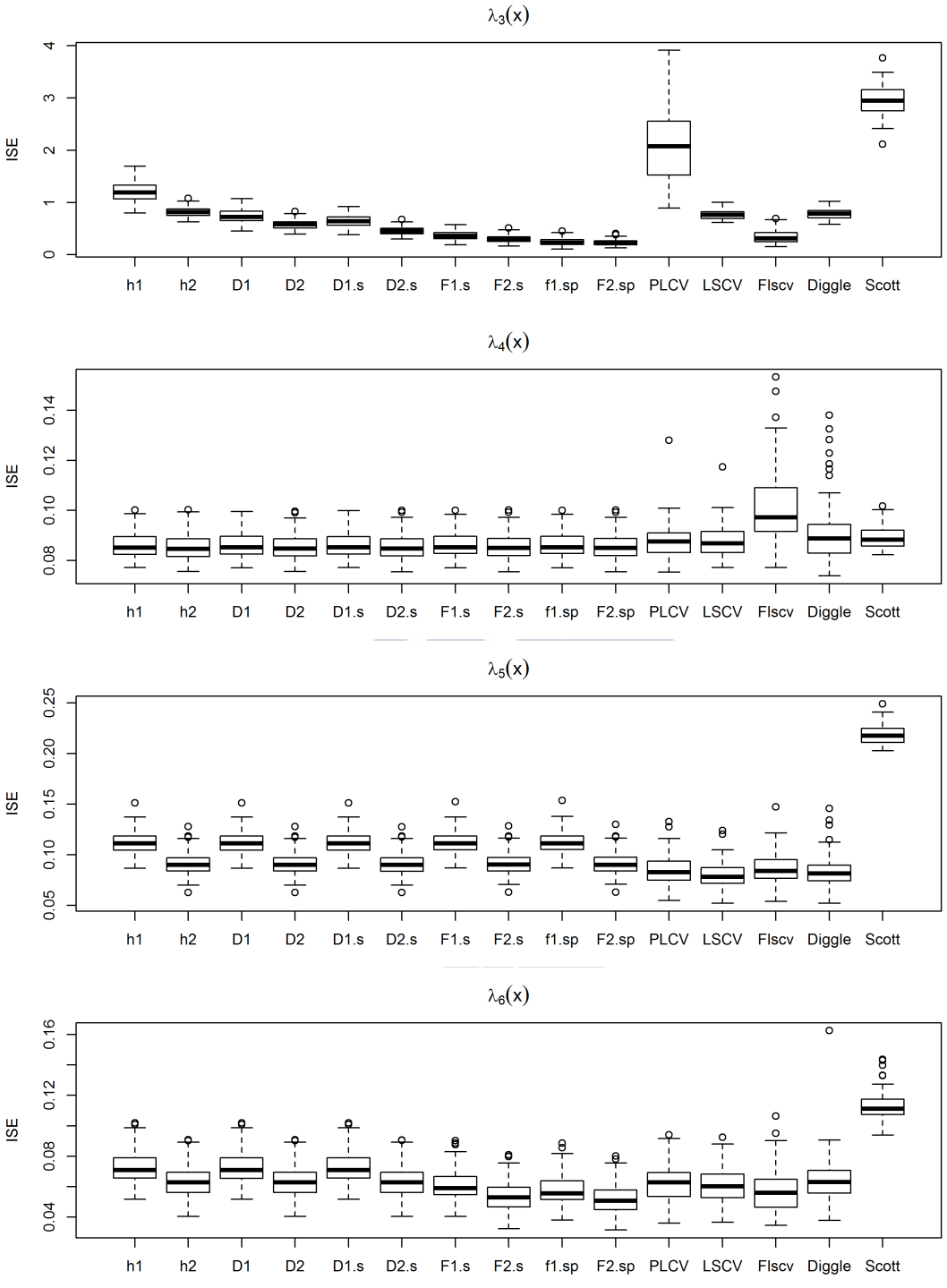


FIGURE 3.5: Comparison of bandwidth selectors for spatial Poisson point patterns with intensities $\lambda_3(x) - \lambda_6(x)$. See the caption of Figure 3.4 for complete information.

The plug-in algorithm is the most affected by the size of the observed patterns. For example, for $m = 500$ 2-stage plug-in and the scalar cross-validation procedures registered similar costs, while for $m = 1000$ 1-stage plug-in showed higher cost than the scalar cross-validation criteria.

If we focus on the plug-in bandwidth selector, we can see the need of a second stage in the most inhomogeneous point patterns (comparison between $\lambda_2(x)$ and $\lambda_3(x)$) and when the first-order intensity can not be represented by a normal reference rule (see $\lambda_5(x)$ and $\lambda_6(x)$). Moreover, diagonal matrices with pre-scaled data and full matrix bandwidths with pre-sphered data outperform their counterparts with untransformed data, except when a scalar bandwidth is sufficient.

For Poisson point patterns with intensity $\lambda_1(x)$ (Table A.2, Figure A.1) the 1-stage plug-in bandwidth selectors performs better than its 2-stage counterpart and we do not find differences between the performance of scalar and matrix bandwidths. The scalar plug-in algorithm and the two cross-validation procedures provide similar results, while Diggle´s criterion leads to higher ISEs. The diagonal plug-in bandwidth and the rule of thumbs lead to similar results. Finally, the full matrix plug-in algorithm outperforms full matrix cross-validation. Both Diggle´s and full matrix cross-validation procedures tend to undersmooth the first-order intensity. For the Thomas cluster process with first-order intensity $\lambda_1(x)$ (Table A.3, Figure A.2) we obtain similar results in the comparison between bandwidth selection procedures, but the performance of the kernel intensity estimator is worse than for the Poisson point process.

For Poisson point patterns with intensity $\lambda_2(x)$ (Table A.4, Figure A.3) our results suggest that we should use a diagonal matrix with pre-scaled data, as it outperforms the scalar bandwidths and provides similar results as the full matrices. We also observe that the plug-in bandwidth selectors outperform the benchmark criteria. In this case, LSCV, PLCV and Scott´s rule of thumbs tend to oversmooth the intensity, while Flscv leads to undersmoothing. As above, we obtain similar results with higher ISEs for the Thomas Cluster processes (Table A.5, Figure A.4).

Comparison between plug-in bandwidth selectors for Poisson point patterns with intensity $\lambda_3(x)$ (Table A.6, Figure A.5) identifies full matrix plug-in with pre-sphered data as the best selector. The 2-stage plug-in scalar bandwidth provides similar results as LSCV and Diggle´s criterion and outperforms PLCV, which tends to oversmooth the first-order intensity. The diagonal plug-in matrix clearly outperforms Scott´s rule of

thumbs, and the full plug-in bandwidth matrix performs slightly better than the cross-validation matrix.

For the log-Gaussian Cox processes with intensity $\lambda_4(x)$ (Table A.7, Figure A.6) the comparison between bandwidth selectors yields results similar to those obtained for the Poisson point processes with first-order intensity $\lambda_1(x)$. However, Figure A6 shows that, in this case, the kernel intensity estimators tend to oversmooth the first-order intensity.

For Poisson point patterns with first-order intensity $\lambda_5(x)$ (Table A.8, Figure A.7), which shows two orthogonal directions of higher intensity, diagonal and full bandwidth matrices do not provide better estimators than the scalar bandwidth. The 2-stage plug-in bandwidth selectors and the benchmark criteria show similar performance, except for the diagonal matrix where the plug-in algorithm clearly outperforms the Scott´s rule of thumbs. Although we do not observe differences in terms of ISE, Figure A.7 shows that Diggle´s criterion and the three cross-validation approaches tend to undersmooth the first-order intensity. For the Poisson point process with intensity $\lambda_6(x)$ (Table A.9, Figure A.8), which has different levels of inhomogeneity in the main directions, the full bandwidth matrix with pre-sphered data outperforms simpler plug-in bandwidths. Comparison between plug-in and the benchmark criteria yields results similar to that obtained for $\lambda_5(x)$, although in this case the plug-in matrix is slightly better than the cross-validation matrix. Figure A.8 shows the same bias towards undersmoothing, as observed for the point patterns with intensity $\lambda_5(x)$, in the bandwidths provided by Diggle´s and the cross-validation criteria.

3.6 Analysis of wildfires registered in Galicia (NW-Spain) during 2006

In this section we analyze the spatial patterns of ignition points for the 5090 arson and the 120 natural wildfires registered in Galicia during the year 2006 (Figure 3.6) in order to compare their spatial structure. 2-stage plug-in bandwidth selection for scalar bandwidth (h2), diagonal matrix with pre-scaled data (D2s) and full matrix with pre-sphered data (F2sp) and the benchmark procedures outlined in Section 3.5 were applied to select the optimal bandwidth and to estimate the first-order intensity functions of both point patterns.

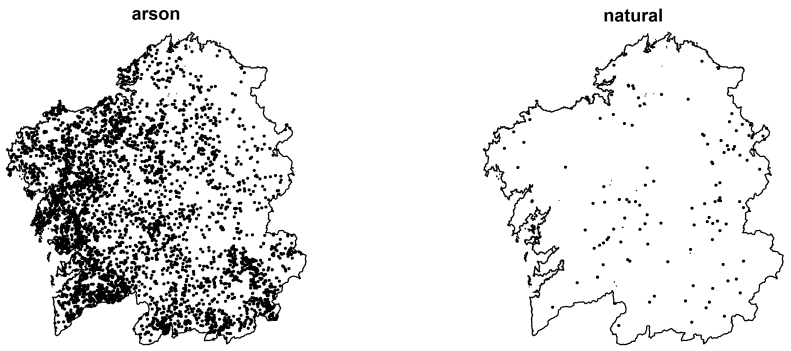


FIGURE 3.6: Ignition points of arson and natural wildfires registered in Galicia during 2006.

Table 3.1 shows the bandwidths provided by each criterion and their computational cost. As observed in the simulation study, the computational demand of the plug-in algorithm increases with the size of the spatial patterns. Diggle´s criterion and the scalar cross-validation procedures were applied considering the boundary represented in Figure 3.6 (map) and with a more regular polygonal boundary (W). We can see that reducing the complexity of the observation region results in a significant reduction in the computational demand and barely affects the selected bandwidth. Particularly, to apply Diggle´s criterion for arson wildfires we need the regular boundary as the map reported a computational error. These results suggest selecting the bandwidth on a regular boundary when the observation region has a highly irregular shape.

Figures 3.7 and 3.8 show the kernel intensity estimators provided by the different bandwidth selectors. For arson fires, the scalar cross-validation (PLCV and LSCV) and Diggle´s criteria lead to a severe undersmoothing of the data, while Scott´s rule of thumbs and the cross-validation matrix provide estimators smoother than those obtained by the plug-in algorithm. The three plug-in bandwidths provide similar estimators, i.e. for this data we could use a scalar bandwidth. For natural wildfires pseudolikelihood cross-validation and the plug-in method with scalar parameter provide similar estimators, as well as for arson fires Scott´s rule and the cross-validation matrix provide smoother estimators, while least-squares cross-validation and Diggle´s criterion lead again to an extreme undersmoothing. Comparison between the plug-in bandwidths shows that the full matrix was able to capture the high intensity in the centre of the region.

Finally comparison between Figures 3.7 and 3.8 shows that during 2006 the spatial distributions of arson and natural wildfires were different. Arson wildfires were mainly concentrated close to the west coast, while the risk of natural wildfires was higher in the east of the region.

		Arson $n = 5090$				Natural $n = 120$			
		h			cost	h			cost
h2		5.61			773.18	15.18			0.14
PLCV	map	2.43			55.40	15.79			48.67
	W	2.35			9.32	14.58			1.89
LSCV	map	0.11			54.79	10.53			48.06
	W	0.11			9.31	9.80			2.03
Diggle	map	NA			NA	8.98			16.80
	W	0.59			0.18	10.70			0.03
		h_1		h_2	cost	h_1		h_2	cost
D2.s		5.61		5.65	774.51	14.33		16.24	0.11
Scott		11.36		11.67	0.02	21.11		21.04	0.00
		H_{11}	H_{21}	H_{22}	cost	H_{11}	H_{21}	H_{22}	cost
F2.sp		32.85	-6.73	35.09	770.91	210.20	37.81	267.29	0.14
Flscv		129.09	-36.17	136.24	41.12	445.53	7.58	442.70	0.08

TABLE 3.1: Optimal bandwidths for arson and natural wildfires and computational costs of the different selectors. For the procedures involving edge-correction, map is the boundary observed in Figure 3.6 and W is a regular boundary containing the observed patterns; h2: scalar plug-in bandwidth; D2.s: Diagonal plug-in bandwidth matrix with pre-scaled data; F2.sp: full plug-in bandwidth matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s bandwidth selector, Scott: Scott´s rule of thumbs.

3.7 Conclusions

In this chapter, we have extended to $\mathbb{R}^2$ the smooth bootstrap method proposed by Cowling et al. (1996) in order to develop a procedure to select the optimal bandwidth matrix for the kernel intensity estimator of inhomogeneous spatial point processes, for which the data driven procedure proposed by Berman and Diggle (1989) for stationary Cox processes has been extensively used.

Under the infill asymptotic framework we have shown that the bootstrap MISE is a consistent estimator of the MISE of the kernel estimator of the density of event locations. This result provides a formal proof of the validity of the smooth bootstrap and suggests the use of the bootstrap AMISE as criterion for bandwidth selection. Therefore, we have proposed a plug-in algorithm based on this measure to select the bandwidth for the kernel intensity estimator. The simulation study shows that plug-in bandwidth selectors

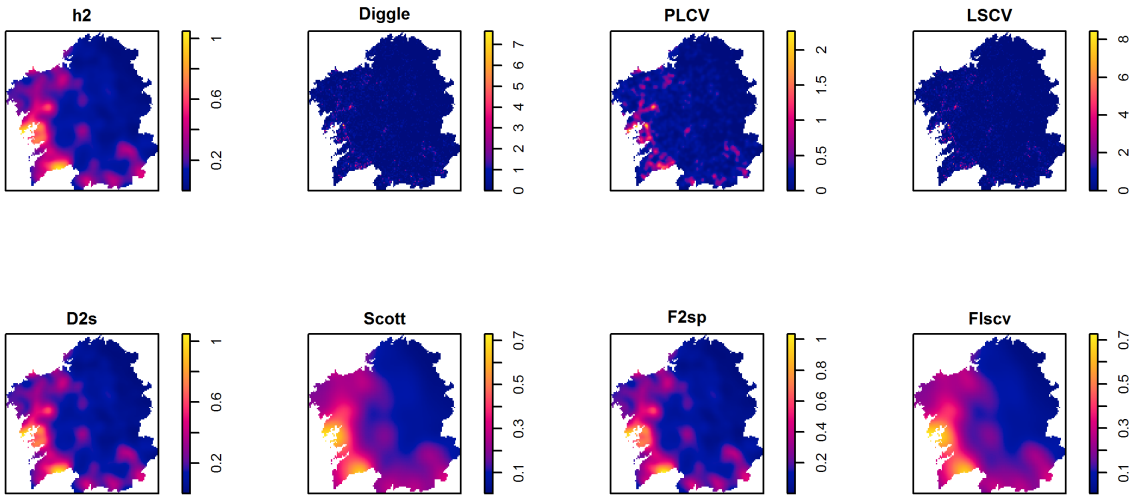


FIGURE 3.7: Kernel intensity estimation for arson wildfires. The label "2" indicates 2-stage plug-in bandwidth selection. h2: scalar bandwidth; D2s: Diagonal matrix with pre-scaled data; F2sp: Full matrix with pre-sphered data. PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s bandwidth selector; Scott: Scott´s rule of thumbs.

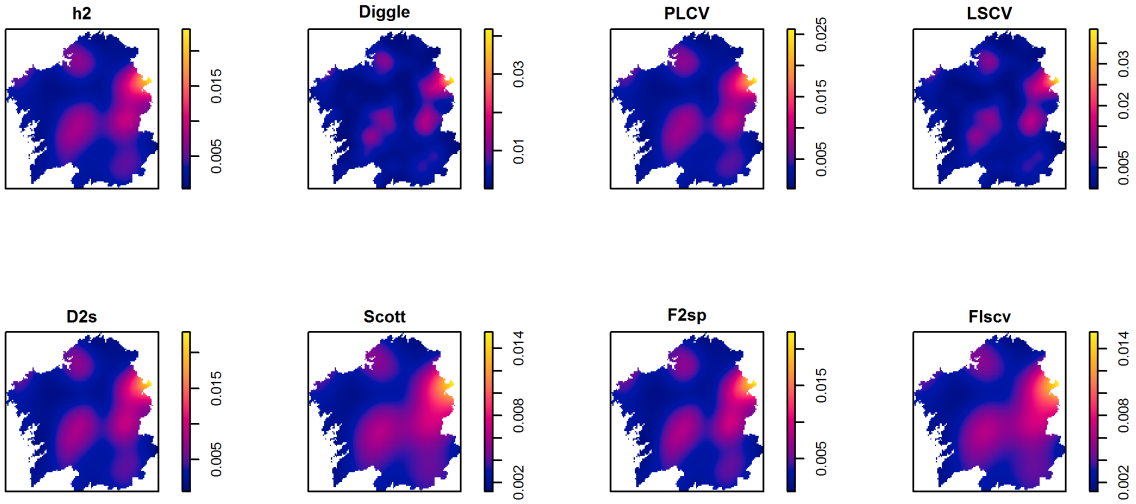


FIGURE 3.8: Kernel intensity estimation for natural wildfires. See detalied information in the caption of Figure 3.7.

are competitive with respect to the criteria currently used for bandwidth selection in spatial point processes for both Poisson and non-Poisson point processes. When we are dealing with highly inhomogeneous data the plug-in algorithm with diagonal and full matrices outperform the benchmark criteria. In the application to real data, we have seen that cross-validation and Diggle´s criterion can lead to a severe undersmoothing of the intensity specially for large point patterns, which confirms that edge-correction can introduce a bias towards undersmoothing, while the plug-in selector provide acceptable estimators. Thus, we can conclude that the plug-in algorithm drives a robust criterion that provides accurate bandwidths for point patterns with different spatial structures and suggest the use of this procedure despite its computational demand. In the application to real data, we can not know a priori whether we need a full matrix or a simpler bandwidth. As the three versions of the plug-in algorithm have similar computational costs, we suggest using the most complex structure.

Finally, once we have proved the consistency of the bootstrap MISE, this performance measure can be used in future works to develop practical and necessary goodness-of-fit tests.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 4

Nonparametric comparison of first-order intensities

4.1 Introduction

A common question in the analysis of multitype spatial point processes is whether two types of events have the same spatial structure. For instance, we can wonder if the spatial distribution of wildfire risk is the same for arson and natural fires. However, up to date the analysis of multitype point processes has been focused on testing for interactions between spatial patterns (see Section 2.3.2). Two spatial patterns with the same structure can be seen as the type i and type j patterns of a random labeled bivariate point process, and consequently $K_i(r) = K_j(r) = K_{ij}(r) = K_j(r)$, which suggest using Monte Carlo tests based on these second-oder properties to answer our question. However differences between the K-function of two inhomogeneous spatial point patterns can reflect differences in the first-order intensities or in the dependence structure. Furthermore, the natural way to test whether two spatial point patterns have the same spatial structures is through the comparison of their first-order intensities.

If two spatial point processes have the same spatial structure, their first-order intensity functions are proportional, and consequently they have the same density of event locations. Therefore, our aim is to develop a nonpaametric test to compare the density of event locations of two inhomogeneous spatial point processes. The equivalence between the density of event locations of spatial Poisson point processes and the bivariate density function of random variables, and the consistency of the kernel estimator of the density of event locations (See Chapter 3), suggest that the test introduced by Duong et al. (2012) for comparison of multivariate distributions may have a good performance

in the point process framework. In Section 4.2 we introduce the test statistic, propose a bootstrap calibration, and discuss the bandwidth selectors needed to implement the test. The performance of the test was checked by a simulation study in Section 4.3 and through its application to the analysis of wildfires in Section 4.4. The chapter ends with some conclusions in Section 4.5.

4.2 Comparison between kernel intensity estimators

A nonparametric test statistic based in the usual squared discrepancy measure is introduced in order to provide a formal comparison of the first-order structure of inhomogeneous spatial Poisson point processes. As explained in Chapter 3, as we cannot distinguish between heterogeneity and interactions on observed patterns without further information, the common practice in the analysis of spatial point processes is to assume that the point process is Poisson, to estimate the first-order intensity, and then to analyze the second-order structure. Therefore, as we are interested on the first-order structure of the observed patterns, in advance we work under the Poisson assumption.

4.2.1 The test statistic

Let $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$ be a realization of a bivariate inhomogeneous spatial Poisson point process observed in a bounded region $W \subset \mathbb{R}^2$, and $\mathbf{X_1} = \{\mathbf{x_1}, \ldots, \mathbf{x_{N_1}}\}$, $\mathbf{X_2} = \{\mathbf{x_{N_1+1}}, \ldots, \mathbf{x_{N_1+N_2}}\}$, where $N = N_1 + N_2$, the spatial patterns of type 1 and type 2 events in $\mathbf{X}$. We denote by $\lambda_1(x)$ and $\lambda_2(x)$ the first-order intensities of $\mathbf{X}_1$ and $\mathbf{X}_2$, and by $\lambda_{01}(x) = \lambda_1(x)/m_1$, $\lambda_{02}(x) = \lambda_2(x)/m_2$ their densities of event locations, where m_1 and m_2 are the expected number of events of each point process. The kernel estimator of the density of event locations for $\mathbf{X}_j$ is

$$\hat{\lambda}_{0j,H_j}(x) = \left(p_{H_j}(x) N_j\right)^{-1} |H_j|^{-1/2} \sum_{i=1}^{N_j} k\left(H_j^{-1/2}\left(x - \mathbf{x_{j,i}}\right)\right) I[N_j \neq 0];\ j = 1, 2 \quad (4.2.1)$$

where the kernel, $k(\cdot)$, is a radially symmetric bivariate probability density function; H_j, is the symmetric and positive-definite bandwidth matrix, and $|H_j|$ denotes the determinant of H_j; and $p_{H_j}(x) = \int_W k\left(H_j^{-1/2}(x - y)\right) dy$ is the edge-correction term.

If $\mathbf{X}_1$ and $\mathbf{X}_2$ have the same spatial distribution, their densities of event locations are equal. Conditional on $N_1 = n_1$ and $N_2 = n_2$, the spatial point patterns can be seen as random samples of the bivariate random distributions with density functions $\lambda_{01}(\cdot)$

and $\lambda_{02}(\cdot)$. Thus, following the proposal of Duong et al. (2012) the null hypothesis, $\mathcal{H}_0 : \lambda_{01}(x) = \lambda_{02}(x)$, can be tested through the squared discrepancy measure

$$T = \int_W \left(\lambda_{01}(x) - \lambda_{02}(x)\right)^2 dx \tag{4.2.2}$$

which can be rewritten as $T = \psi_{0,1} + \psi_{0,2} - (\psi_{0,12} + \psi_{0,21})$, where $\psi_{0,j} = \int_W \lambda_{0j}(x)^2 dx$ and $\psi_{0,ij} = \int_W \lambda_{0i}(x)\,\lambda_{0j}(x)\,dx$, for $j = 1, 2$. Overlooking edge-effects we obtain the following test statistic

$$\hat{T} = \hat{\psi}_{0,1} + \hat{\psi}_{0,2} - \left(\hat{\psi}_{0,12} + \hat{\psi}_{0,21}\right) \tag{4.2.3}$$

where

$$\hat{\psi}_{0,1} = \frac{1}{n_1^2} \sum_{i_1=1}^{n_1} \sum_{i_2=1}^{n_1} k_{G_1}\left(\mathbf{x_{i_1}} - \mathbf{x_{i_2}}\right)$$

$$\hat{\psi}_{0,2} = \frac{1}{n_2^2} \sum_{j_1=n_1+1}^{n} \sum_{j_2=n_1+1}^{n} k_{G_2}\left(\mathbf{x_{j_1}} - \mathbf{x_{j_2}}\right)$$

$$\hat{\psi}_{0,12} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} k_{G_1}\left(\mathbf{x_i} - \mathbf{x_j}\right)$$

$$\hat{\psi}_{0,21} = \frac{1}{n_1 n_2} \sum_{i=1}^{n_1} \sum_{j=n_1+1}^{n} k_{G_2}\left(\mathbf{x_i} - \mathbf{x_j}\right)$$

G_1 and G_2 are the bandwidth matrices for the kernel estimators of the density functionals $\psi_{0,1}$ and $\psi_{0,2}$. Intuitively, the test statistic $\hat{T}$ compares pairwise distances between events within the same pattern $(\psi_{0,1}, \psi_{0,2})$ with pairwise distances between events from different patterns $(\psi_{0,12}, \psi_{0,21})$. When $\mathbf{X_1}$ and $\mathbf{X_2}$ have different spatial structures, the distances between events from different patterns are expected to be larger than those between events within the same pattern.

In the multivariate density framework Duong et al. (2012) obtained the asymptotic distribution of the test statistic $\hat{T}$. If all terms in G_j tend to 0 and $n_j|G_j|^{1/2} \to \infty$ as $n_j \to \infty$ for $j = 1, 2$; $f_j, j = 1, 2$ have bounded and square integrable partial derivatives up to order 2, and $k(\cdot)$ is a continuous, symmetric, square integrable density function such that $\int_{\mathbb{R}^2} uu^T k(u)du = \mu_2(k)I_2$, with $\mu_2(k) < \infty$, then

$$Z = \frac{T - \mu_T}{\sigma_T \sqrt{1/n_1 + 1/n_2}} \to N(0, 1) \tag{4.2.4}$$

where $\mu_T = \left(n_1^{-1}|G_1|^{-1/2} + n_2^{-1}|G_2|^{-1/2}\right)$ and $\sigma_T^2 = 3\int f(x)^3 dx - \left(\int f(x)^2 dx\right)^2$.

To use the asymptotic null distribution we first need to estimate μ_T and σ_T. Duong et al. (2012) used the algorithm developed by Chacón and Duong (2010) to estimate the bandwidth matrices G_1 and G_2 minimizing the asymptotic mean square error of $\hat{\psi}_1$ and $\hat{\psi}_2$ to estimate μ_T. The variance of $\hat{T}$ is estimated by $\hat{\sigma}_T = (n_1\hat{\sigma}_1 + n_2\hat{\sigma}_2)\,/\,(n_1+n_2)$, where $\hat{\sigma}_j = \left[D\hat{f}_{j,G_j}(\bar{X}_j)\right]^T S_j \left[D\hat{f}_{j,G_j}(\bar{X}_j)\right]$, being $D\hat{f}_{j,G}(x) = n_j^{-1}\sum_{i=1}^{n_j} Dk_G(x-\mathbf{x_i})$ the kernel estimator for the gradient of f_j, $\bar{X}_j$ the sample mean and S_j the sample variances of $\mathbf{X_j}$, and $G_j = \left(2n_j^{-2}/3\right)^{1/8} S_j$ the normal scale bandwidth selector for the kernel estimator of the density derivatives.

The equivalence between the density of event locations of inhomogeneous spatial Poisson point processes and the density function of bivariate random values would allow us to establish the asymptotic normality of $\hat{T}$ in the point processes framework as a direct extension of expression (4.2.4). However, the asymptotic normal calibration of the test statistic provided by property (4.2.4) can produce a test that does not achieve the nominal significance level, specially when we compare small point patterns, or realizations of non-Poisson point processes. For this reason, we use a smooth bootstrap procedure (see Section 3.3) to generate resamples of $\mathbf{X}_1$ and $\mathbf{X_2}$ under $\mathcal{H}_0$ and estimate the null distribution of $\hat{T}$. The bootstrap test can be implemented as follows:

1. Let $\mathbf{X_1}$ and $\mathbf{X_2}$ be the observed patterns, and $\mathbf{X} = \{\mathbf{X_1}, \mathbf{X_2}\}$ the corresponding bivariate spatial point pattern. Obtain the kernel intensity estimator of the unmarked pattern $\mathbf{X}$, $\hat{\lambda}_H(x)$.
2. Compute the test statistic $\hat{T}$.
3. For $b = 1, \ldots, B$:
   - 3.1 Generate a bivariate spatial point process $\mathbf{X_b^*} = \{\mathbf{X_{1,b}^*}, \mathbf{X_{2,b}^*}\}$ where for $j = 1, 2$, $\mathbf{X_{j,b}^*}$ are realizations of spatial Poisson point processes with first-order intensities $n_j\hat{\lambda}_H(x)$, being n_j the number of event in $\mathbf{X_j}$.
   - 3.2 Compute the test statistic $\hat{T}_b^*$.
4. The relative position of $\hat{T}$ in the ordered sample $\hat{T}_{(b)}^*,\ b = 1, \ldots, B$ provides the probability of rejecting the null hypothesis.

4.2.2 Bandwidth selectors

To implement the bootstrap test introduced above we need to conduct two bandwidth selection procedures. The bandwidth matrix for the kernel intensity estimator in step

1 can be obtained by the plug-in bandwidth selector introduced in Section 3.4.2. We also need bandwidth matrices $G_j, j = 1, 2$ for the kernel estimators of the squared density integrals $\psi_{0,j}, j = 1, 2$ in (4.2.3) to conduct steps 2 and 3.2. To estimate these bandwidths we use the plug-in procedure proposed by Chacón and Duong (2010). The AMSE-optimal bandwidth matrix for $\hat{\psi}_{0,G}$ is that minimizing its asymptotic square bias

$$AB^2(G) = \| \, n^{-1}|G|^{-1/2} + \frac{\mu_2(k)}{2} \sum_{i=1}^{2} G_{ii}\psi_{2e_i} \, \|^2 \tag{4.2.5}$$

To obtain the bandwidth minimizing $AB^2(G)$ we need pilot estimators of $\psi_{2e_i}, i = 1, 2$, which also require bandwidth selection. Thus considering scalar pilot bandwidths, $G_r = g_r I_2$, to ease computation, and using the same notation as in Chapter 3, the optimal $|r|$th-order pilot bandwidths can be defined as those minimizing expression (3.3.7)

$$AMSE\left(\hat{\psi}_{r,g_r}\right) = 2n^{-2}g_r^{-2|r|-2}\psi_0 R(k^{(r)}) + \left[n^{-1}g_r^{-|r|-2}k^{(r)}(0) + \frac{g_r^2}{2}\mu_2(k)\sum_{i=1}^{2}\psi_{r+2e_i}\right]^2$$

which are

$$g_{r,AMSE} = \left[\frac{-2k^{(r)}(0)}{\mu_2(k)\left(\sum_{i=1}^{2}\psi_{r+2e_i}\right)}\right]^{-1/(r+4)} \tag{4.2.6}$$

In order to overcome the limitation of scalar pilot bandwidths we can apply a spheric transformation to the event locations prior to conduct the plug-in algorithm. Therefore, following the same ideas as in Section (3.4.2), G_{AMSE} can be estimated throughout a l-stage plug-in algorithm

1. Given a spatial pattern $\mathbf{X}$, let S be the covariance matrix of the event locations, obtain the pre-sphered data $\{\mathbf{x}^*{}_i\}_{i=1}^n = \{S^{-1/2}\mathbf{x_i}\}_{i=1}^n$.

2. Let $j_{max} = 2l + 4$. Obtain the normal reference estimators $\hat{\psi}_{j_{max}}^{NR}$ for the pre-sphered pattern $\mathbf{X}^*$. Plug these estimators into expression 4.2.6 to obtain the j_{max}th AMSE optimal bandwidths.

3. For $j = j_{max} - 2, \ldots, 4$ and X^*.

   3.1 Calculate kernel estimators of ψ_r of order $j = |r|$ using the plug-in bandwidths $g_{r,AMSE}$.

   3.2 Plug $\hat{\psi}_{r,g_{r,AMSE}}$ into equation (4.2.6) to obtain plug-in estimates of order $j-2$.

4. Use $g_{2e_i,AMSE}$ to obtain $\hat{\psi}_{2e_i,g_{2e_i,AMSE}}$. Plug these estimates into equation (4.2.5).

5. Minimize (4.2.5) to obtain the optimal bandwidth matrix G^*_{AMSE} for $\hat{\psi}_{0,G^*}$. To this purpose we can apply the Newton-Raphson numerical algorithm.

6. Back-transform G^*_{AMSE} into the original scale by $G_{AMSE} = S^{1/2} G^*_{AMSE} S^{1/2}$.

In practice, 2-stage plug-in bandwidth selectors are commonly used, as they provide a good balance between computational cost and accuracy.

4.3 Simulation study

We have conducted a simulation study to analyze the performance of the nonparametric test based on $\hat{T}$. We have checked the normality of $\hat{T}$ under the null hypothesis, and compared the probabilities of rejecting $\mathcal{H}_0$ provided by the normal and bootstrap calibrations. In addition we have analyzed the power of the test for different ways and degrees of departure from the null hypothesis.

We first generated inhomogeneous spatial Poisson point processes (IPP1) with first-order intensity functions $\lambda_j(x) = a_j \exp(-3x_2), j = 1, 2$, for any $x = (x_1, x_2) \in \mathbb{R}^2$ on the unit square. Different values of $a_j, j = 1, 2$ were considered to obtain different proportions of type 1 and type 2 events in order to check whether the asymmetry in the size of the point patterns affects the performance of the test. The alternative hypothesis was generated adding events on a subregion of the type 2 point pattern as follows

$$\begin{cases} \lambda_1(x) = a_1 \exp(-3y) \\ \lambda_2(x) = a_2 \left(\exp(-3y) + \dfrac{100}{1700} \epsilon \left(\dfrac{1}{4} \right)^2 I_{(x_1, x_2 \in [0.5, 0.75])} \right) \end{cases} \tag{4.3.1}$$

Using $a_1 = a_2 = 1700$ we generate realizations of a bivariate point process with $m_1 = m_2 = 500$. $(1/4)^2$ in expression (4.3.1) is the area of the subregion of the unit square where $100(1/4)^2\epsilon$ events are added to generate realizations of the alternative hypothesis in the balanced design, and ϵ, which takes values between 0.2 and 0.8 at 0.2 intervals, determines the number of events added to the type 2 point process, that is the level of departure from the null hypothesis (Figure 4.1). These processes are an inhomogeneous and discretized version of the spatio-temporal point processes used in Díaz-Avalos et al. (2013) to check the performance of nonparametric separability tests.

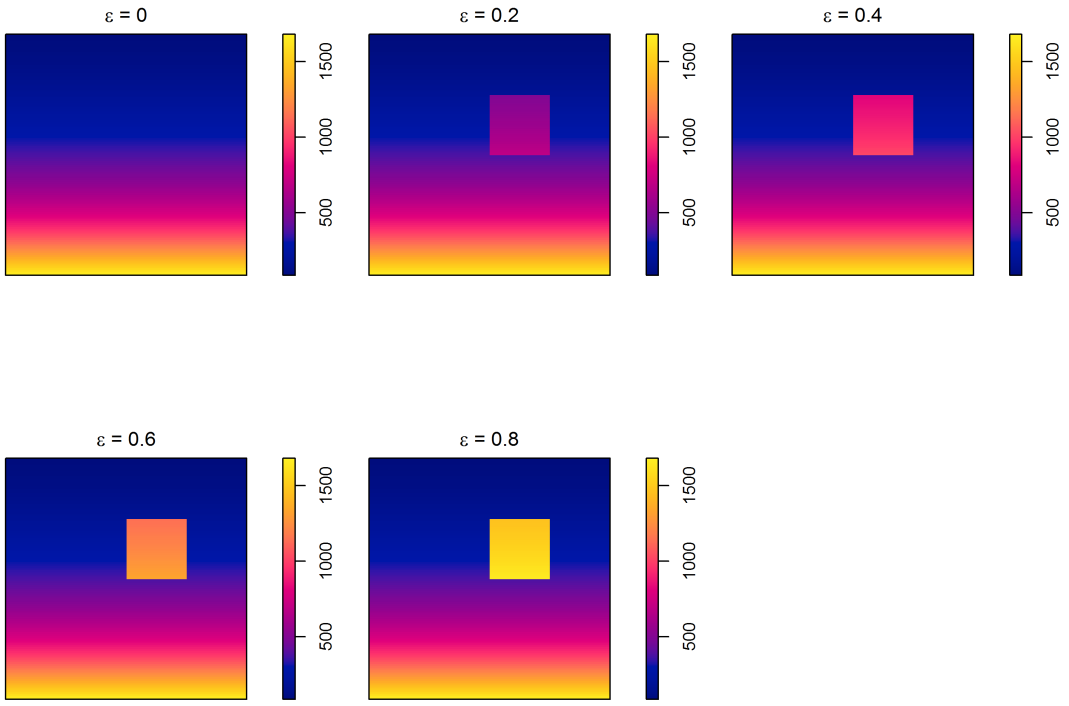


FIGURE 4.1: First-order intensities for the type 1 and type 2 evens in the bivariate point processes with marginal intensity functions (4.3.1) and balanced design.

We also tested the performance of the test for pairwise comparisons on a multitype inhomogeneous spatial Poisson point process (IPP2) with marginal intensities

$$\lambda_j(x) = a_j \phi\left(\epsilon - 0.2y, 0.02\right), \; j = 1, \ldots, 9, \; \epsilon = 0, \ldots, 0.1 \tag{4.3.2}$$

where $\phi(\mu, \sigma)$ denotes the density function of the univariate normal distribution with mean μ and variance σ^2, and a_j ranked from 500 to 50 at 50 intervals to generate marginal patterns with different sizes. As we can see in Figure (4.2) the high-intensity area moves upwards as ϵ increases.

We generated 1000 realizations of the different scenarios defined by the values of ϵ in expressions (4.3.1) and (4.3.2), and conducted the asymptotic and bootstrap tests based on $\hat{T}$. The bootstrap distribution of $\hat{T}$ under $\mathcal{H}_0$ was obtained from $B = 1000$ realizations of the null hypothesis ($\epsilon = 0$). The probabilities of rejecting $\mathcal{H}_0$ at any significance level, α, were computed as the proportion of times the p-value provided by each calibration was smaller than α. The simulation study was conducted with the help of the *spatstat* (Baddeley and Turner, 2005) and *ks* (Duong, 2013a) packages of R (R Core Team, 2014).

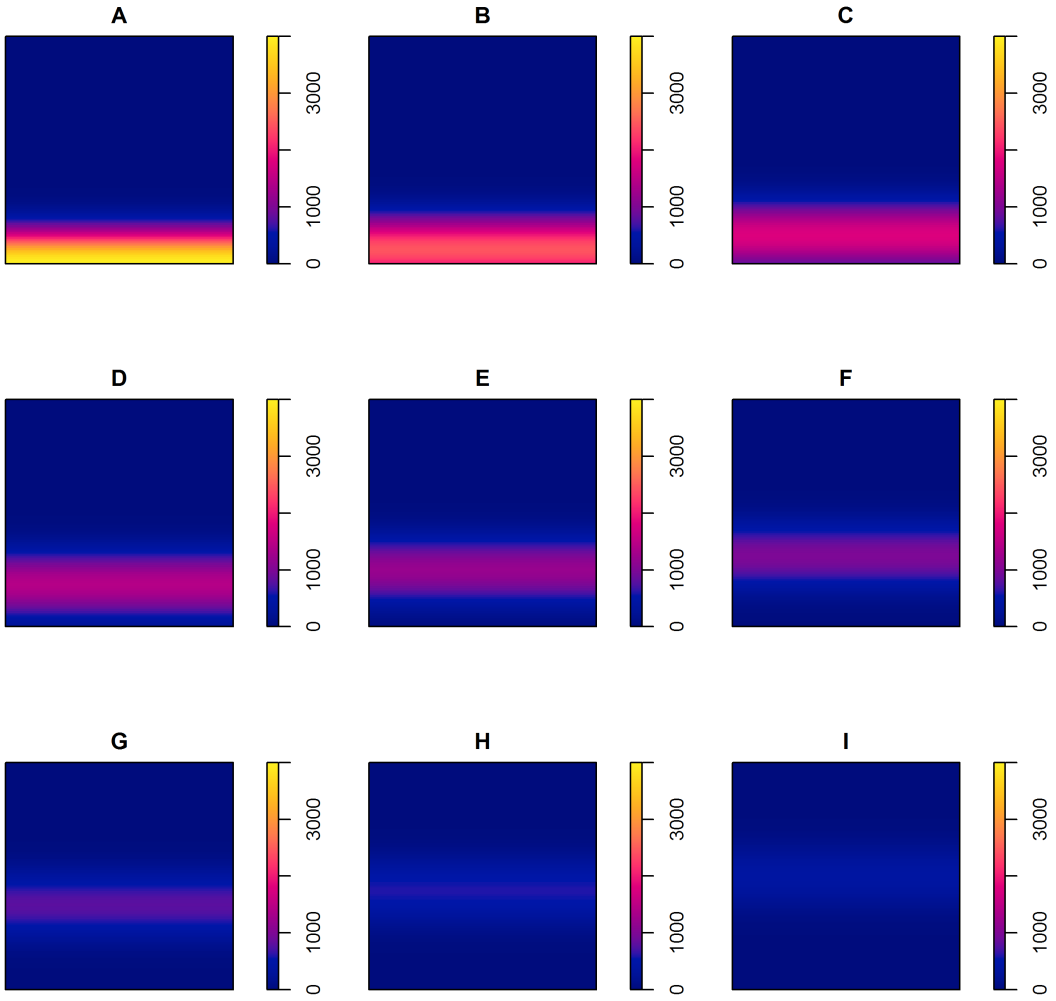


FIGURE 4.2: Multitype point process: marginal intensities of the type A-I point processes with first-order intensity functions (4.3.2).

For the point processes with univariate intensities defined in expression (4.3.1), Figure 4.3 and Table 4.1 show that $\hat{T}$ is not normal distributed. In addition the normal calibration overestimates the variance of $\hat{T}$ given rise to a conservative test. The bootstrap calibration provides type I errors close to the nominal significance level, although the test can be slightly conservative for unbalanced designs. As expected, given their behavior under $\hat{H}_0$ the bootstrap test is more powerful than the asymptotic test. The power of the test increases with the departure from the null hypothesis, and decreases for unbalanced designs.

Table 4.2 and Figure 4.4 show that the null distribution of $\hat{T}$ for pairwise comparisons

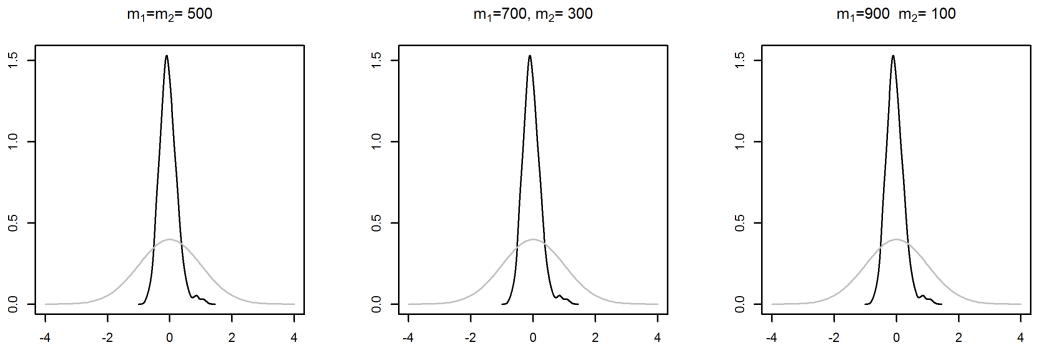


FIGURE 4.3: Comparison between the empirical distribution of the standardized test statistic, Z, for IPP1 (black) and the standard normal distribution (grey)

			$\alpha = 0.05$		$\alpha = 0.01$		$\hat{Z}$		
	ϵ	N_{2_ϵ}	A	B	A	B	$\hat{\mu}_Z$	$\hat{\sigma}_Z$	shapiro-test
$m_1 = 500$	0		< 0.001	0.048	¡ 0.001	0.007	-0.05	0.29	$<$ 1e-9
$m_2 = 500$	0.2	20	< 0.001	0.103	< 0.001	0.022			
	0.4	40	0.015	0.315	0.002	0.134			
	0.6	60	0.263	0.778	0.067	0.587			
	0.8	80	0.780	0.994	0.465	0.978			
$m_1 = 700$	0		< 0.001	0.044	< 0.001	0.009	-0.04	0.30	$<$ 1e-9
$m_2 = 300$	0.2	12	0.001	0.064	0.000	0.010			
	0.4	24	0.017	0.263	0.002	0.124			
	0.6	36	0.137	0.718	0.032	0.517			
	0.8	48	0.471	0.888	0.187	0.705			
$m_1 = 900$	0		< 0.001	0.035	< 0.001	0.004	-0.07	0.35	$<$ 1e-9
$m_2 = 100$	0.2	4	0.004	0.081	< 0.001	0.025			
	0.4	8	0.011	0.080	0.001	0.022			
	0.6	12	0.050	0.242	0.013	0.058			
	0.8	16	0.129	0.466	0.037	0.177			

TABLE 4.1: Performance of the asymptotic (A) and bootstrap (B) tests based on $\hat{T}$ for IPP1. Probabilities of rejection with the asymptotic (A) and bootstrap (B) distributions of $\hat{T}$ at nominal significance levels $\alpha = 0.05$ and $\alpha = 0.01$. Empirical mean $\hat{\mu}_Z$, and standard deviations $\hat{\sigma}_Z$, and p-value of the Shapiro-Wilk test for the standardized test statistic under $\mathcal{H}_0$ ($\hat{Z}$). For the alternative hypothesis ϵ determines the degree of departure from $\mathcal{H}_0$ and $N_{2,eps}$ the expected number of events added to the original pattern.

in IPP2 (4.1) are not normal. Table 4.2 also shows that the bias of $\hat{\mu}_T$ increases as the number of events in each pattern decreases. In contrast, $\hat{\sigma}_Z$ approximates 1 as the pattern sizes decrease, leading to more conservative tests for large patterns. The bootstrap calibration is less affected by the sample sizes and provides type I errors reasonably close to the nominal significance levels. Although, for type H events the nominal calibration performs better than the bootstrap calibration for $\alpha = 0.01$, the general behavior of the later under the null hypothesis is better.

Tables 4.4 and 4.3 show that the bootstrap test is more powerful than the asymptotic test. For both calibrations, the power increases with the size of the observed patterns. As for IPP1, the probability of rejecting the null hypothesis increases with the discrepancy between the univariate first-order intensities, in fact we only obtain probabilities of rejection lower than 1 for consecutive patterns.

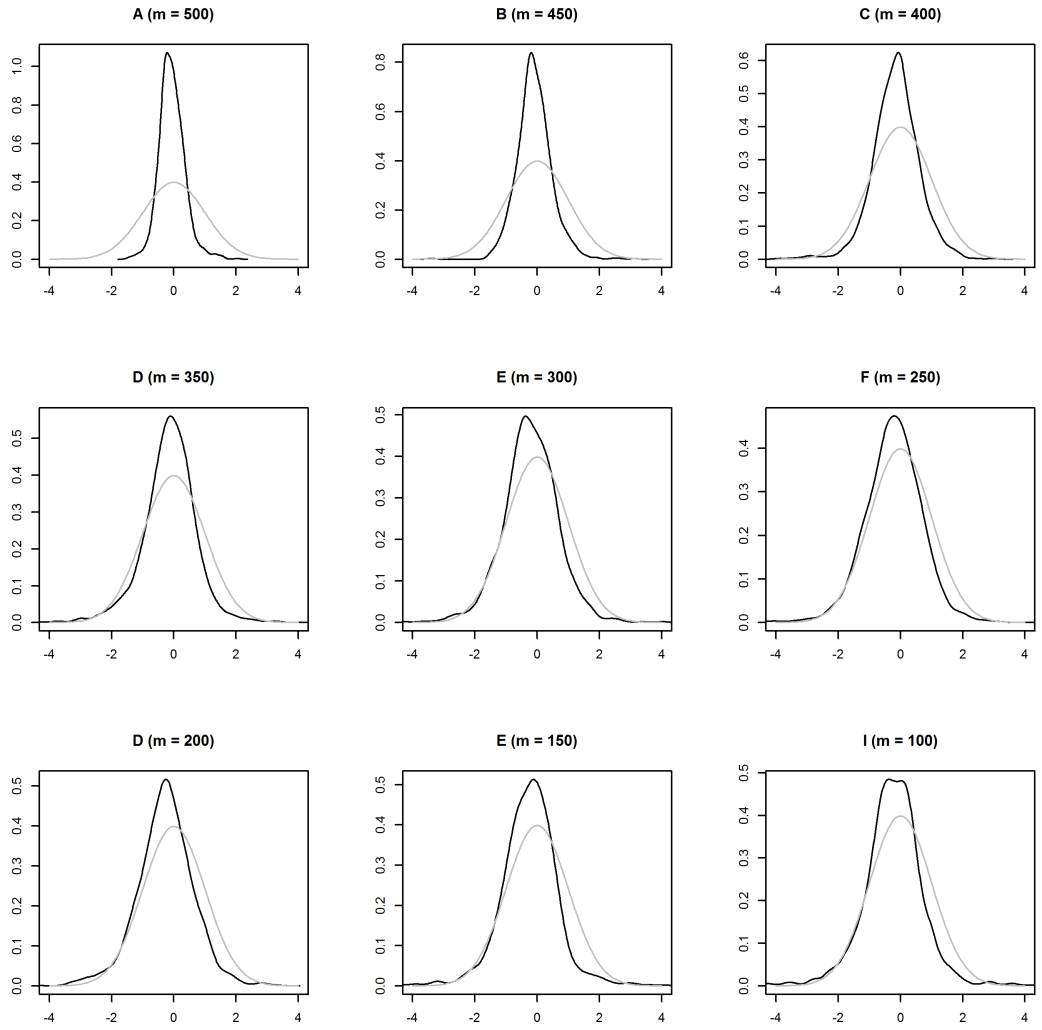


FIGURE 4.4: Comparison between the empirical distribution of the standardized test statistic, Z, for the IPP (black) and the standard normal distribution (grey)

		$\alpha = 0.05$		$\alpha = 0.01$		$\hat{Z}$		
	m_j	A	B	A	B	$\hat{\mu}_Z$	$\hat{\sigma}_Z$	shapiro-test
A	500	0.001	0.067	< 0.001	0.016	-0.06	0.42	$<$ 1e-9
B	450	0.005	0.044	0.003	0.013	-0.09	0.56	$<$ 1e-9
C	400	0.012	0.037	0.002	0.004	-0.16	0.77	$<$ 1e-9
D	350	0.020	0.052	0.008	0.009	-0.15	0.88	$<$ 1e-9
E	300	0.018	0.059	0.007	0.014	-0.23	0.92	$<$ 1e-9
F	250	0.022	0.045	0.007	0.014	-0.24	1.00	$<$ 1e-9
G	200	0.021	0.088	0.005	0.016	-0.28	0.92	$<$ 1e-9
H	150	0.025	0.045	0.009	0.006	-0.27	0.96	$<$ 1e-9
I	100	0.023	0.062	0.010	0.014	-0.24	0.96	$<$ 1e-9

TABLE 4.2: Performance of the test under the null hypothesis for multitype inhomogeneous Poisson point processes (IPP2). Comparison between empirical p-values for asymptotic (A) and bootstrap (B) distributions of $\hat{T}$, empirical mean,$\hat{\mu}_Z$ and standard deviations $\hat{\sigma}_Z$ for the standardized test statistic,and p-value of the Shapiro-Wilk test.

		Asymptotic test								
$\alpha = 0.05$		A	B	C	D	E	F	G	H	I
Bootstrap test	A		0.56	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	B	0.93		0.86	1.00	1.00	1.00	1.00	1.00	1.00
	C	1.00	0.95		0.95	1.00	1.00	1.00	1.00	1.00
	D	1.00	1.00	0.98		0.97	1.00	1.00	1.00	1.00
	E	1.00	1.00	1.00	1.00		0.92	1.00	1.00	1.00
	F	1.00	1.00	1.00	1.00	0.96		0.93	1.00	1.00
	G	1.00	1.00	1.00	1.00	1.00	0.85		0.75	1.00
	H	1.00	1.00	1.00	1.00	1.00	1.00	0.80		0.55
	I	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.67	

TABLE 4.3: Power of the T-test at significance level $\alpha = 0.05$, for 1000 realizations of multitype point processes (IPP2). Comparison between asymptotic (upper triangular) and bootstrap (lower triangular) tests.

		Asymptotic test								
$\alpha = 0.01$		A	B	C	D	E	F	G	H	I
Bootstrap test	A		0.27	1.00	1.00	1.00	1.00	1.00	1.00	1.00
	B	0.81		0.65	1.00	1.00	1.00	1.00	1.00	1.00
	C	1.00	0.82		0.87	1.00	1.00	1.00	1.00	1.00
	D	1.00	1.00	0.92		0.88	1.00	1.00	1.00	1.00
	E	1.00	1.00	1.00	0.96		0.81	1.00	1.00	1.00
	F	1.00	1.00	1.00	1.00	0.87		0.70	1.00	1.00
	G	1.00	1.00	1.00	1.00	1.00	0.82		0.55	1.00
	H	1.00	1.00	1.00	1.00	1.00	1.00	0.64		0.35
	I	1.00	1.00	1.00	1.00	1.00	1.00	1.00	0.52	

TABLE 4.4: Power of the T-test at significance level $\alpha = 0.01$, for 1000 realizations of multitype point processes (IPP2). Comparison between asymptotic (upper triangular) and bootstrap (lower triangular) tests.

4.4 Comparison between spatial patterns of wildfires in Galicia

In this Section we show the applicability of the nonparametric test introduced in this chapter to check whether the spatial distribution of wildfires registered in Galicia depends on their burned area or cause. To this purpose, we have considered the spatial pattern of the 6904 wildfires registered in Galicia during 2006 classified according to burned area: 4800 small fires ($S < 1$ ha), 1767 regular fires ($1 \leq S < 25$ ha), and 337 large wildfires ($S \geq 25$ ha) and by cause: 5090 arson, 120 natural, 400 due to negligence, 475 reproductions, and 819 with unknown cause.

Pairwise comparisons between the first-order intensity functions of wildfires according to burned area (Figures 4.5 confirmed that the three point patterns have different spatial distribution (p-values < 0.001 with both asymptotic and bootstrap calibrations). Pairwise comparison between wildfires pattern by cause (Figure 4.6) also detected significant differences between their first-order intensities (p-value < 0.001 with both asymptotic and bootstrap calibrations (Figure 4.6).

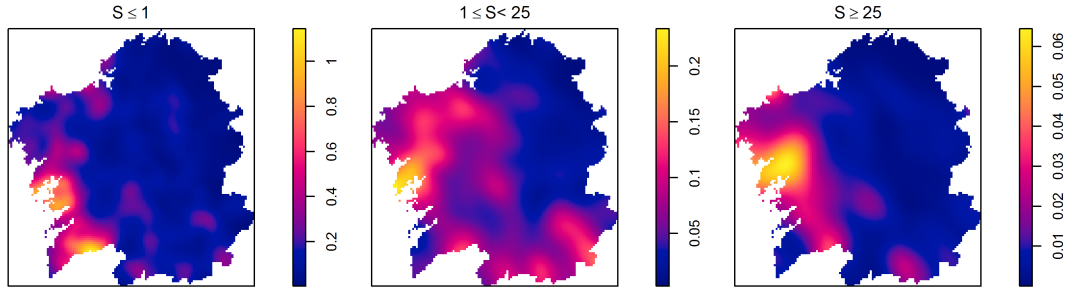


FIGURE 4.5: Kernel intensity estimator of the wildfires registered in 2006 by burned area (different scales)

The test statistic $\hat{T}$ can also be used to test the separability assumption on spatio-temporal point processes with discrete temporal component. For instance, we can stratify the spatio-temporal pattern of wildfires registered in 2006 by months and test whether their spatial structure change over time. The number of wildfires shows an important seasonal variability, with maximum values in summer (2230 fires in August) and minimum values in autumn (6 ignition points in October). We have considered the wildfires

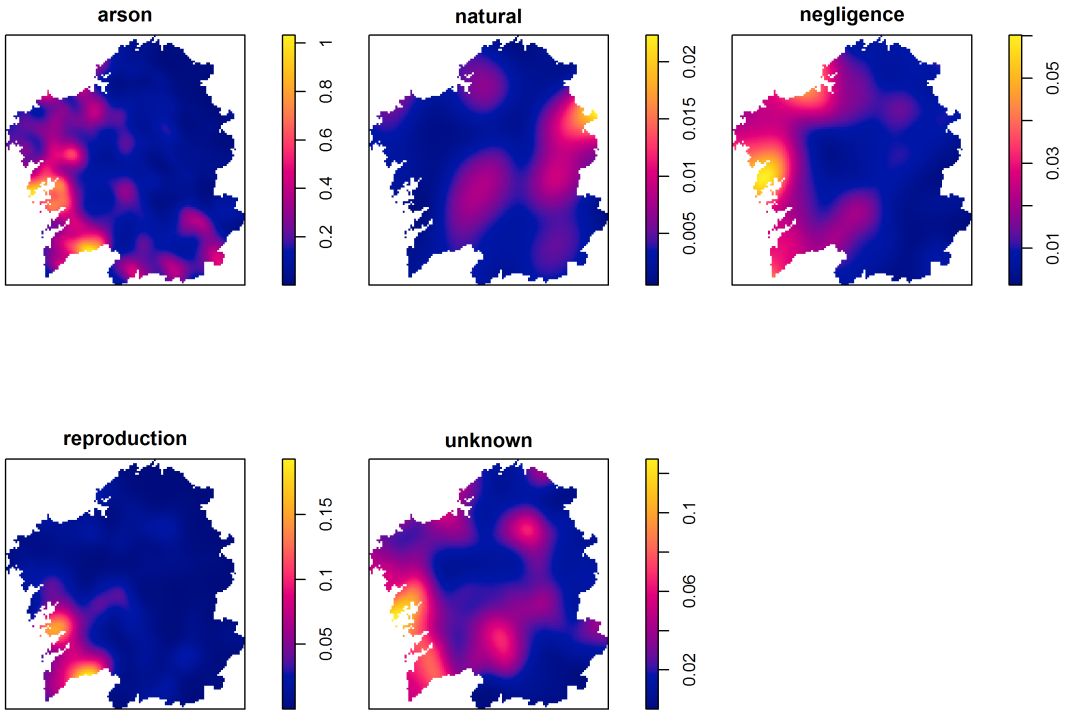


FIGURE 4.6: Kernel intensity estimator of the wildfires registered in 2006 by cause (different scales)

registered between January and September, as the small number of events registered during the last 3 months do not allow estimating the corresponding first-order intensity functions.

The kernel intensity estimators in Figure 4.7 show that the region with high incidence of wildfires shifted from the south area of Galicia, from January to March, to the west coast, which registered the highest fire intensity during summer. The asymptotic and bootstrap tests (4.5) confirmed the temporal variability in the spatial patterns of ignition points, being April and May the only pair of month with homogeneous spatial structure.

4.5 Conclusions

In this chapter, following the idea Duong et al. (2012) for comparison of d-variate random distributions, we propose a nonparametric test baesd on square differences between

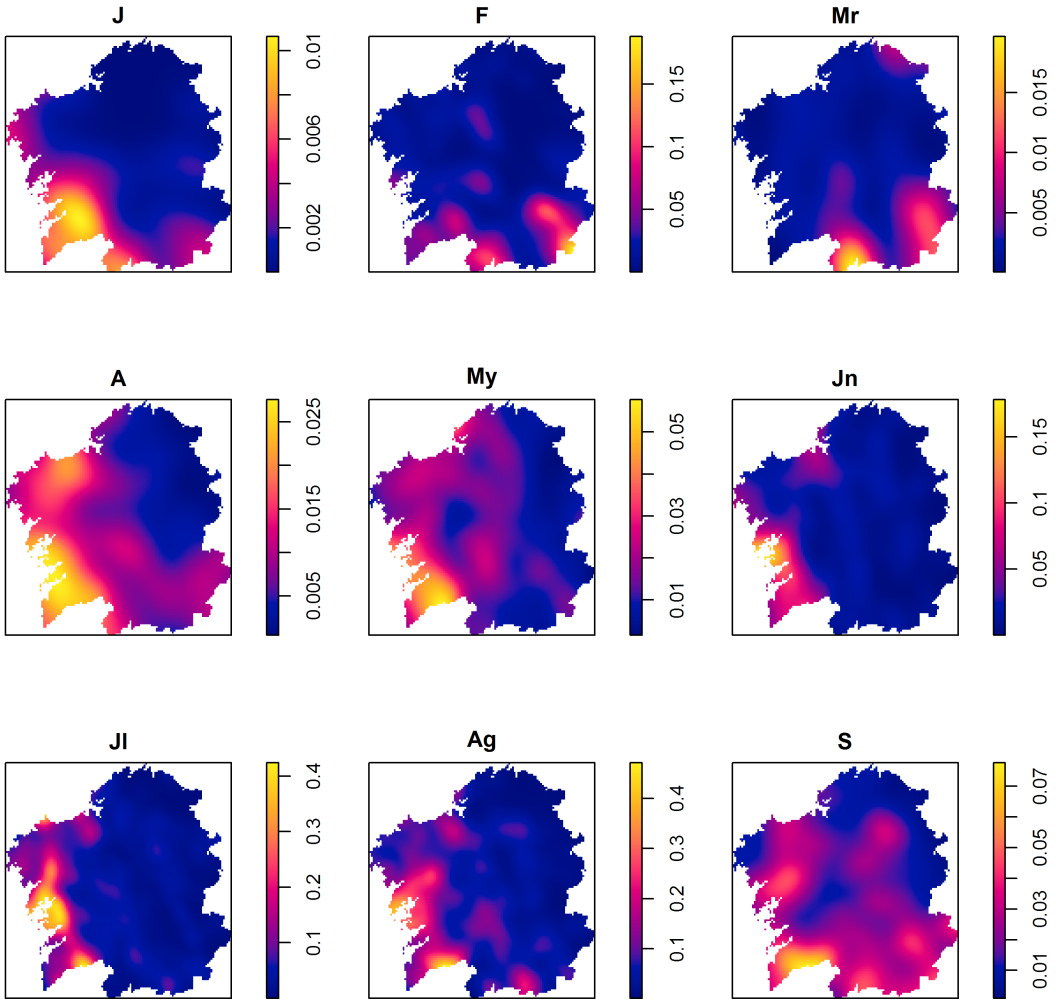


FIGURE 4.7: First-order intensity functions for the spatial patterns of wildfire registered in Galicia from January to September 2006 (different scales)

density of event locations to compare the first-order structure of inhomogeneous spatial point patterns. The relationship between the density of event locations of spatial Poisson point processes and the bivariate density function, as well as the asymptotic properties of the kernel estimator of the density of event locations and the kernel estimators of the integrated density derivatives shown in Chapter 3, suggest that our test statistic may inherit the asymptotic normal distribution obtained in the multivariate framework. However, the results obtained in the simulation study lead to reject the normality of the null distribution of $\hat{T}$ for inhomogeneous spatial Poisson point processes. Given the lack of normality of the test statistic, we have proposed a bootstrap calibration procedure.

To implement the test we use kernel estimators of the first-order intensities and the integrated square density functions to obtain the test statistic and to conduct the bootstrap

			Asymptotic test								
		n	J	F	Mr	A	My	Jn	Jl	Ag	S
			60	659	89	260	441	638	1789	2238	669
Bootstrap test	J	60		0	0	0.025	0	0.007	0.001	0	0
	F	659	0		0	0	0	0	0	0	0
	Mr	89	0	0.014		0	0	0	0	0	0
	A	268	0	0	0		0.215	0	0	0	0
	My	441	0	0	0	0.09		0	0	0	0
	Jn	638	0	0	0	0	0		0	0	0
	Jl	1789	0	0	0	0	0	0		0	0
	Ag	2238	0	0	0	0	0	0	0		0
	S	669	0	0	0	0	0	0	0	0	

TABLE 4.5: P-values of the pairwise comparison of the spatial distribution of wildfires registered in Galicia during 2006. Asymptotic calibration in the upper triangular and bootstrap calibration in the lower triangular (0 indicate p-value < 0.001) . We considered the period January-September as less than 30 fires by month were observed from October onwards. n is the number of wildfires registered each month.

procedure. We suggest using 2-stage plug-in algorithms with scalar pilot bandwidths to select the optimal bandwidth matrices for both kernel estimators.

The simulation study shows that the asymptotic test can be quite conservative. The bootstrap procedure provided reasonable calibrations, although its performance can be affected by differences in the size of the spatial point patterns under comparison. Our results indicate that the bootstrap test is more powerful than the asymptotic test. The test was able to detect different ways of departure from the null hypothesis, and its power increases with the departure from $\mathcal{H}_0$. However the power of the test reduces for small point patterns, and for unbalanced designs.

In the application to the analysis of wildfires the asymptotic and bootstrap test provided similar results. However, in view of the simulation study we recommend using the bootstrap calibration. We have also seen that this test can be used as spatio-temporal separability test if we stratify the temporal component into subgroups. Arranging the temporal component into subgroups is recommended when data are collected at coarse level.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 5

Analysis of the spatial patterns of wildfires in Galicia

5.1 Introduction

This chapter is an extension of the spatial analysis of wildfires recorded in the Fonsagrada-Ancares district conducted in Fuentes-Santos et al. (2013). We have conducted a preliminary analysis of the wildfires registered in Galicia during the period 1999-2008 using the statistical techniques currently available for spatial point processes (see Chapter 2, as well as the new techniques introduced in Chapters 3 and 4.

The first step in the analysis of any observed pattern is testing the CSR assumption, which implies stationarity and independent events (Section 2.2.1). Figures 1.6-1.9 suggest that the spatial distribution of wildfires in Galicia is not homogeneous. Thus, as indicated in Section 2.2.6.2, the analysis of the wildfire patterns was divided in two steps: in Section 5.2 we estimate the first-order intensities of the observed patterns, test the stationarity assumption, and compare the spatial distribution of different wildfire patterns; in Section 5.3 we analyze the interactions between wildfires within the same pattern and the interactions between different fire patterns.

5.2 Spatial structure of wildfires

5.2.1 Stationarity test and first-order intensity estimates

We applied the stationarity test introduced by Comas et al. (2009) (see Section 2.2.6.2) to test whether the risk of wildfires depend on the spatial location. For this purpose, the kernel intensity estimator in (2.2.15) was obtained using the bandwidth matrix provided by the 2-stages plug-in algorithm proposed in Section 3.4.2.

Tables 5.1-5.3 report the plug-in bandwidth matrices for each wildfire pattern, and the result of the stationarity test. We show the value of the test statistic (2.2.15) for the observed pattern, $\hat{S}_{obs}$, and the maximum value of the test statistic for 19 Monte Carlo simulations of homogeneous Poisson processes, S^*_{max}, which provides a test with significance level $\alpha = 0.05$. The stationarity test rejects the null hypothesis for the whole datasets, and for the spatial patterns of wildfires classified by year (Table 5.1). Figure 5.1 shows that the region of higher fire incidence varied over time, although the southern area of Pontevedra, in the boundary with Portugal, registered high fire intensities during the whole period.

		Bandwidth			**Stationarity test**	
	n	H_{11}	H_{12}	H_{22}	$\hat{S}_{obs}$	S^*_{max}
$1999-2008$	85127	8.87	-0.356	6.872	55556.64	4548.92
1999	8561	37.297	-0.558	27.572	5688.74	1412.42
2000	12497	27.062	-1.631	20.136	8397.74	1700.3
2001	9924	28.156	-3.619	23.23	7344.95	1497.32
2002	10714	24.769	-2.261	20.817	7687.41	1590.43
2003	8513	33.098	0.773	29.031	5175.86	1412.91
2004	10587	25.245	-1.388	21.564	7798.47	1574.33
2005	11908	24.77	0.285	19.793	9197.27	1664.99
2006	6904	28.482	-3.031	30.632	4779.41	1272.21
2007	3053	43.659	-11.503	45.302	2461.42	845.05
2008	2466	45.105	-12.271	47.888	2110.05	739.67

TABLE 5.1: Stationarity test for the wildfire patterns by year. n is the number of wildfires; H, the 2-stages plug-in bandwidth matrix; $\hat{S}_{obs}$, the value of the discrepancy measure 2.2.15 for the observed pattern, and S^*_{max} the maximum of the discrepancy measure for 19 Monte Carlo simulations of homogeneous Poisson point processes.

The stationarity assumption was also rejected for the spatial patterns of wildfires classified by burned area (Table 5.2). In Figure 5.2, which shows the first-order intensity of large wildfires by years, we see that the southern area of Ourense was among the regions with high incidence of large fires all years, except during 2006 when large fires shifted

			Bandwidth			Stationarity test	
Year	Size	n	H_{11}	H_{12}	H_{22}	$\hat{S}_{obs}$	S^*_{max}
	$S < 1$	6607	36.938	-1.195	31.310	3869.19	650.65
1999	$1 \leq S < 25$	1900	63.843	-0.964	49.588	1361.65	273.80
	$S \geq 25$	54	293.661	34.714	242.616	43.78	23.65
	$S < 1$	8881	29.995	-1.144	24.214	5493.85	816.69
2000	$1 \leq S < 25$	3357	42.751	-7.733	36.998	2333.56	427.27
	$S \geq 25$	259	169.558	4.179	111.194	236.66	73.36
	$S < 1$	7246	31.024	-3.558	27.246	4893.85	680.50
2001	$1 \leq S < 25$	2582	52.346	-12.995	43.650	1876.19	350.51
	$S \geq 25$	96	412.302	-90.696	165.636	81.91	33.45
	$S < 1$	7579	25.677	-1.616	21.952	5415.82	801.91
2002	$1 \leq S < 25$	2977	59.867	-15.035	57.682	1674.71	359.11
	$S \geq 25$	158	163.106	-59.232	193.538	129.01	56.69
	$S < 1$	6235	34.611	2.439	30.646	3461.67	620.08
2003	$1 \leq S < 25$	2145	67.350	-13.239	74.846	953.97	300.84
	$S \geq 25$	133	179.095	36.997	236.156	99.61	42.31
	$S < 1$	7732	26.230	-0.950	22.407	5507.64	807.35
2004	$1 \leq S < 25$	2657	58.202	-15.345	56.370	1529.78	325.84
	$S \geq 25$	198	227.026	-57.909	216.051	119.92	57.88
	$S < 1$	8722	25.404	1.091	21.105	6414.72	877.69
2005	$1 \leq S < 25$	2886	64.584	-15.046	49.507	1806.10	359.03
	$S \geq 25$	300	272.425	-83.386	201.285	200.18	64.73
	$S < 1$	4800	30.182	-1.576	32.930	3156.81	587.34
2006	$1 \leq S < 25$	1767	79.796	-17.614	83.946	892.22	220.58
	$S \geq 25$	337	112.980	-35.728	121.552	270.94	92.79
	$S < 1$	2312	48.849	-9.089	51.900	1578.02	309.71
2007	$1 \leq S < 25$	694	69.810	-26.742	77.584	574.05	148.90
	$S \geq 25$	47	283.995	-91.248	202.815	46.83	25.81
	$S < 1$	1744	59.853	-11.035	62.416	1280.97	250.49
2008	$1 \leq S < 25$	685	69.181	-31.060	76.113	557.46	164.02
	$S \leq 25$	37	429.217	-58.789	306.802	33.72	23.91

TABLE 5.2: Stationarity test for the wildfire patterns by burned area. n is the number of wildfires; H, the 2-stages plug-in bandwidth matrix; $\hat{S}_{obs}$, the value of the discrepancy measure 2.2.15 for the observed pattern, and S^*_{max} the maximum of the discrepancy measure for 19 Monte Carlo simulations of homogeneous spatial Poisson point processes.

to the west coast of Galicia.

For wildfires classified by cause, the stationarity test accepted the null hypothesis for the spatial patterns of natural wildfires during 2000, and reproductions during 2008 (5.3). However these results should be taken with caution given the small size of both spatial patterns ($n < 30$). Figures 5.3 and 5.5 show that the spatial patterns of arson fires were similar to those observer for the total amount of fires by year, while natural fires were more frequent in the east area of Galicia. Appendix B.1 provides the first-order

intensities of wildfires classified by size and cause for the 10 years under study.

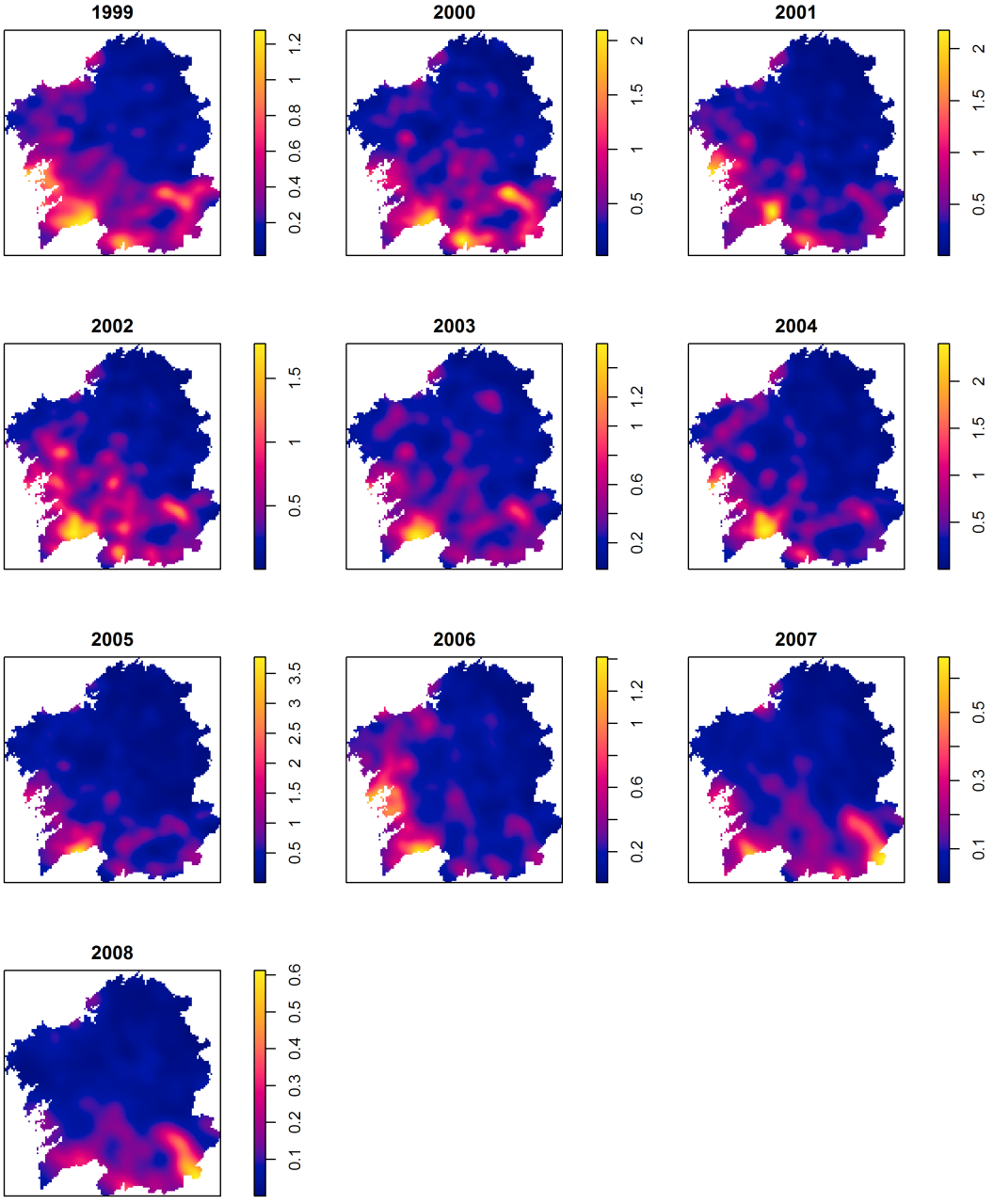


FIGURE 5.1: First-order intensity estimator for the spatial patterns of wildfires by year (different scales)

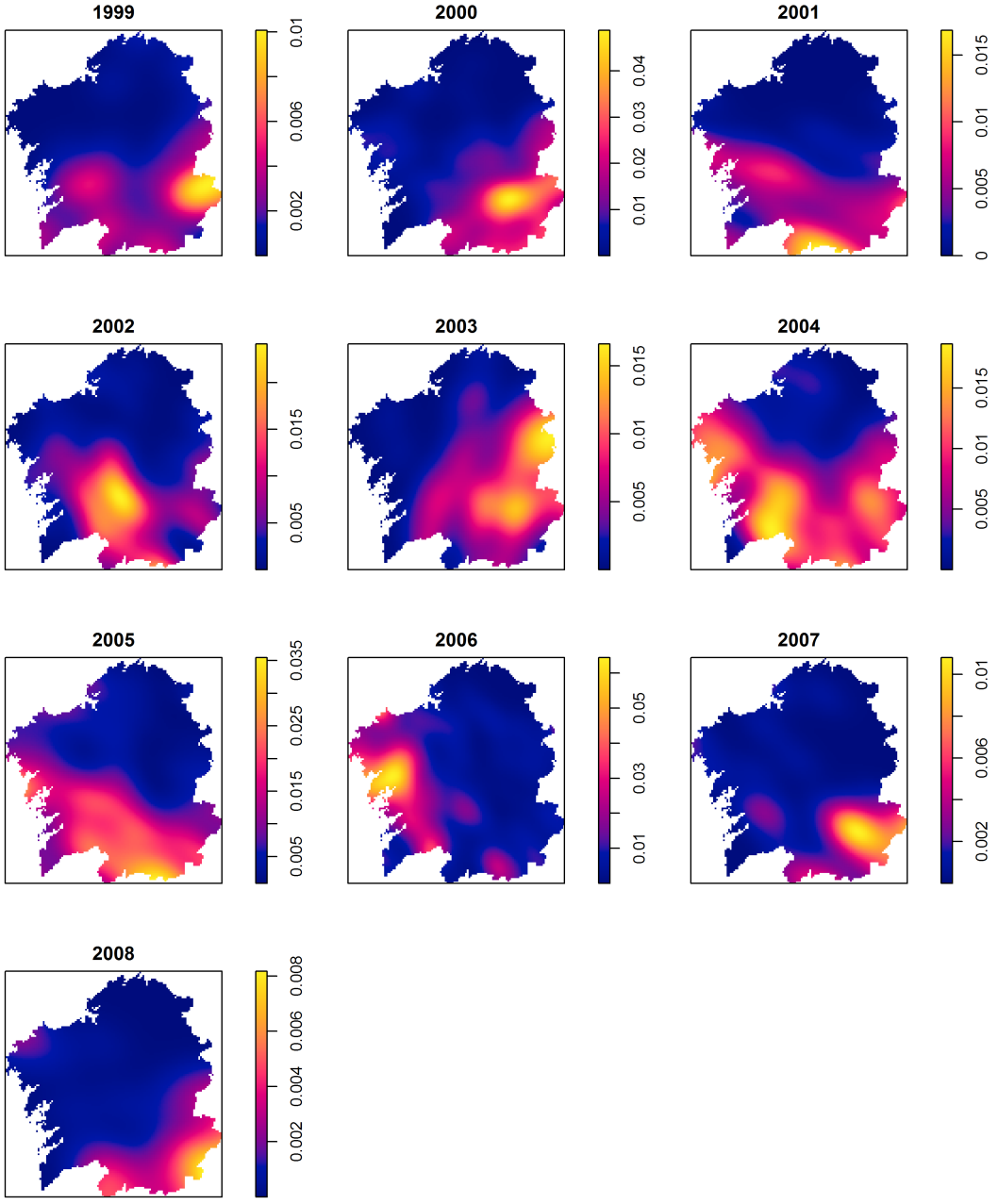


FIGURE 5.2: First-order intensity estimator for the spatial patterns of large wildfires by year (different scales)

			Bandwidth			**Stationarity test**	
Year	Cause	n	H_{11}	H_{12}	H_{22}	$\hat{S}_{obs}$	S^*_{max}
	arson	7020	40.190	-0.802	28.215	4348.41	644.74
	natural	96	240.429	83.226	267.504	87.05	32.63
1999	negligence	563	137.664	2.612	182.796	247.37	98.47
	reproduction	142	222.451	35.085	228.077	86.84	44.93
	unknown	740	45.884	9.767	70.503	661.68	171.04
	arson	10846	27.593	-1.627	20.361	6788.11	966.63
	natural	16	637.812	28.029	542.866	7.12	12.44
2000	negligence	573	101.540	6.740	132.080	328.98	111.65
	reproduction	232	159.750	-8.399	61.158	224.78	73.06
	unknown	830	46.604	12.051	62.236	702.91	187.52
	arson	8576	28.989	-4.459	22.667	6052.99	843.93
	natural	94	260.648	14.457	238.245	56.97	28.51
2001	negligence	433	99.039	8.268	193.182	257.04	98.59
	reproduction	162	149.012	48.378	112.528	158.96	63.91
	unknown	659	46.046	9.386	86.185	603.35	148.33
	arson	9208	27.705	-3.465	21.751	5988.71	868.19
	natural	72	133.861	2.192	318.063	58.75	31.93
2002	negligence	400	150.453	-9.325	199.404	201.00	81.85
	reproduction	319	72.871	2.778	38.600	389.91	119.11
	unknown	715	37.209	3.255	61.155	742.25	180.31
	arson	6949	33.715	-0.469	30.349	3599.55	685.22
	natural	223	109.577	16.704	150.371	165.90	73.60
2003	negligence	336	117.780	28.432	179.140	214.21	76.81
	reproduction	288	123.850	14.776	98.504	220.51	82.79
	unknown	717	58.192	6.524	80.872	558.05	153.12
	arson	8881	26.544	-2.647	22.571	5947.59	864.87
	natural	102	140.276	36.121	216.579	79.26	42.04
2004	negligence	416	117.951	15.229	170.651	263.12	88.55
	reproduction	284	66.320	9.953	53.699	342.98	105.79
	unknown	904	43.953	-1.370	69.036	855.88	197.44
	arson	9694	28.182	-0.935	22.600	6874.25	884.92
	natural	72	146.312	92.412	271.425	66.18	34.75
2005	negligence	519	103.560	-2.703	138.290	339.18	109.58
	reproduction	564	54.913	6.930	38.851	653.08	163.60
	unknown	1059	54.778	2.910	54.049	945.47	206.63
	arson	5090	32.846	-6.734	35.092	3410.25	572.77
	natural	120	210.197	37.810	267.287	58.15	46.09
2006	negligence	400	130.701	12.312	164.045	232.88	80.73
	reproduction	475	59.020	1.884	68.923	449.72	128.37
	unknown	819	83.351	1.731	101.014	445.55	153.92
	arson	2147	47.269	-14.755	46.554	1612.77	354.88
	natural	48	286.464	155.709	657.427	20.82	17.66
2007	negligence	323	134.248	-4.326	216.885	180.14	75.46
	reproduction	80	345.076	-42.037	298.707	52.51	29.71
	unknown	455	155.899	-3.184	119.528	299.21	90.95
	arson	1806	44.188	-13.801	45.541	1498.50	299.15
	natural	44	381.192	10.006	493.312	24.69	19.82
2008	negligence	261	234.249	-2.472	299.098	111.01	49.79
	reproduction	28	559.943	-84.966	787.225	15.54	16.17
	unknown	327	120.406	-10.656	116.466	233.65	85.23

TABLE 5.3: Stationarity test for the wildfire patterns by cause. See details in the caption of Table 5.1.

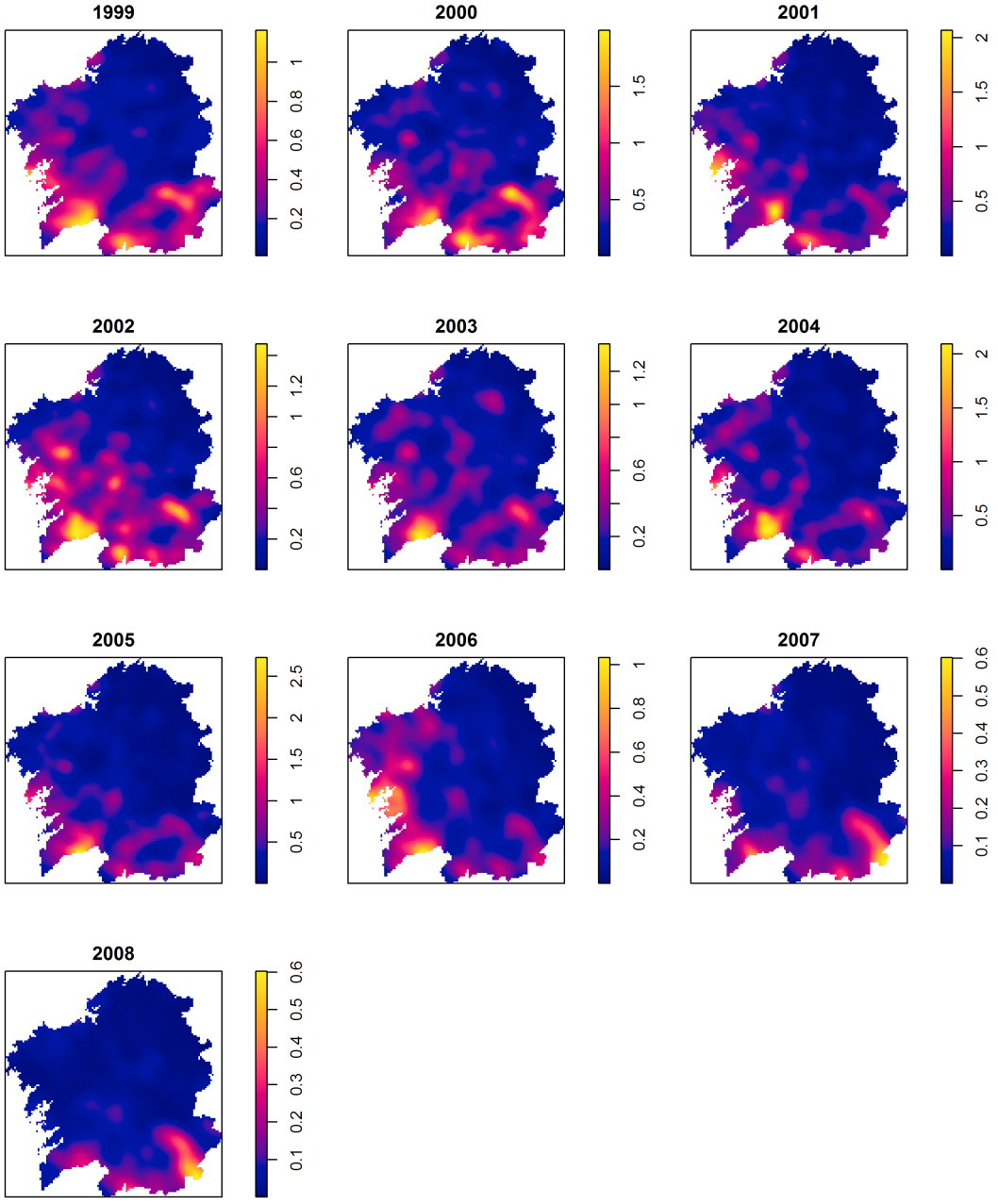


FIGURE 5.3: First-order intensity estimator for the spatial patterns of arson wildfires by year (different scales)

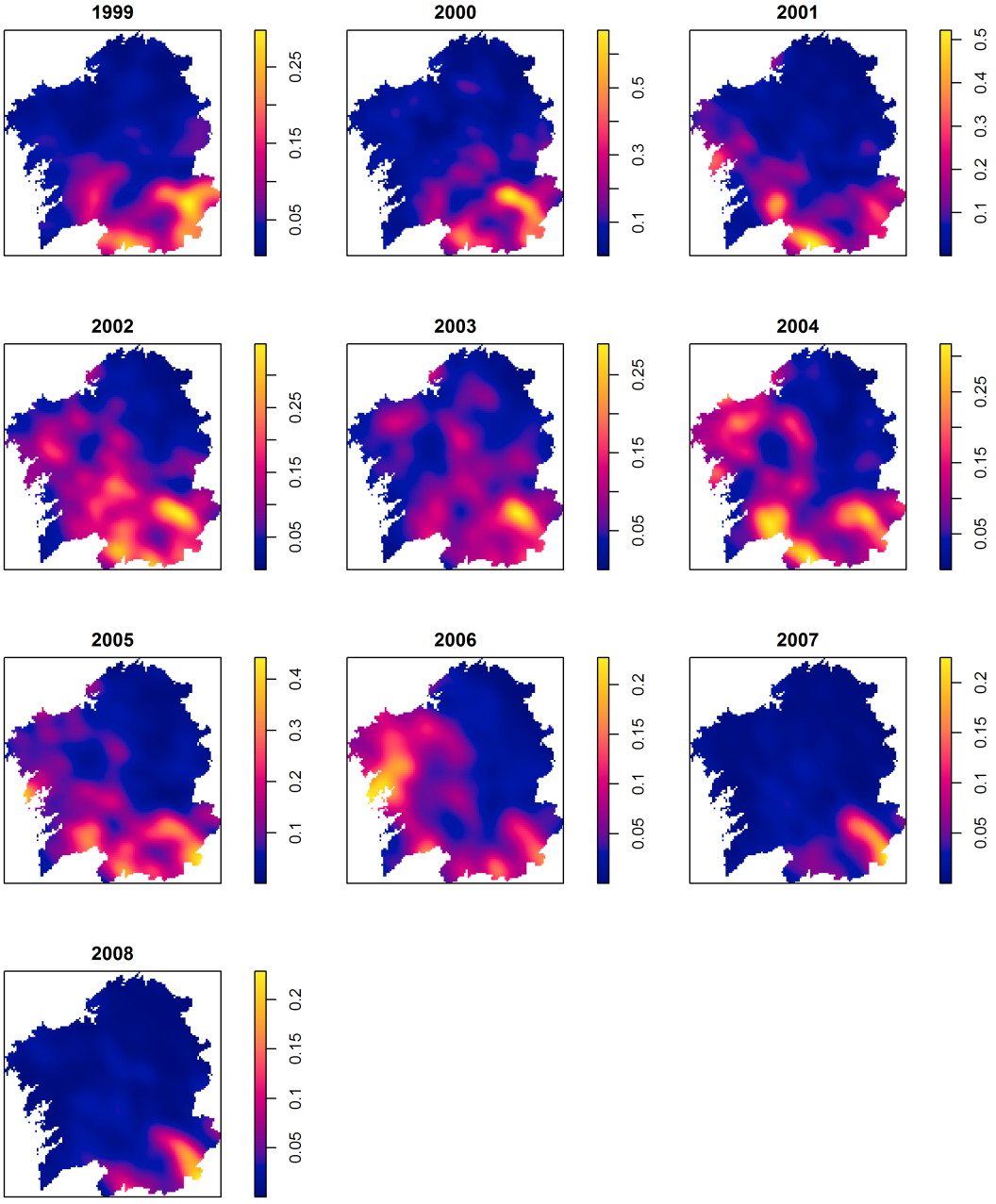


FIGURE 5.4: First-order intensity estimator for the spatial patterns of arson wildfires with burned area higher than 1 km by year (different scales)

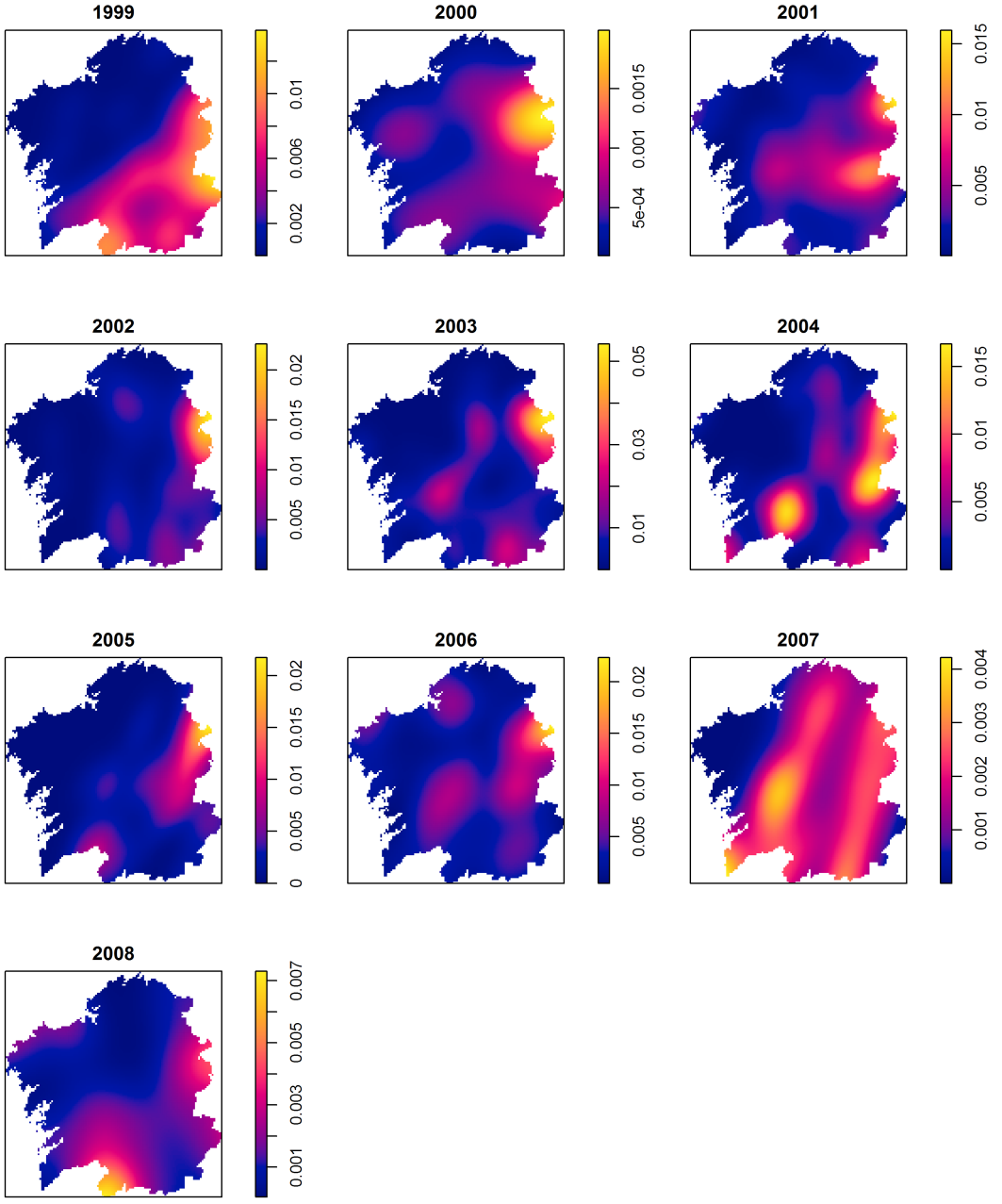


FIGURE 5.5: First-order intensity estimator for the spatial patterns of arson wildfires by year (different scales)

5.2.2 Comparison of wildfire patterns

The first-order intensity estimators shown in Figures 5.2-5.5 suggest that the spatial distribution of large, natural and arson wildfires varies over time. In this section we have used the nonparametric test with bootstrap calibration introduced in Chapter 4 to analyze the spatial variability of these types of wildfire over years.

The nonparametric test detected significant differences between years for the spatial distribution of large fires (Table 5.4) except for the pair 2000-2007. Figure 5.2 shows that large fires in 2000 and 2007 where mainly concentrated in the east area of Ourense. Interannual differences in the spatial patterns of large fires is expected, as a region affected by fires needs some years to recover the vegetation, which is the combustible for a new fire.

Table 5.5 show significant differences in the spatial distributions of natural fires between years, except for the pairs 1999-2008, and 2006-2007. The test did not detect differences between 2000 and the remainder years, but these results should be seen with caution given the low number of natural fires registered in 2000.

Given the large number of arson fires, and the subsequent computational cost of the bootstrap calibration, we have compared the first-order intensities of arson fires with more than 1 ha burned. The nonparametric test detected differences between years (p-value < 0.001), i.e. interannual variability in the spatial distribution of arson fires (Figure 5.4).

	n	1999	2000	2001	2002	2003	2004	2005	2006	2007
1999	54									
2000	259	0.012								
2001	96	0.004	<0.001							
2002	158	0.002	<0.001	<0.001						
2003	133	<0.001	<0.001	<0.001	<0.001					
2004	198	<0.001	<0.001	<0.001	<0.001	<0.001				
2005	300	<0.001	<0.001	0.048	<0.001	<0.001	0.02			
2006	337	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001		
2007	47	0.002	0.264	<0.001	<0.001	<0.001	<0.001	<0.001	<0.001	
2008	37	0.008	0.022	0.026	<0.001	<0.001	0.002	0.002	<0.001	0.034

TABLE 5.4: P-values of the pairwise comparison of the spatial distribution of large wildfires between years.

	n	1999	2000	2001	2002	2003	2004	2005	2006	2007
1999	96									
2000	16	0.518								
2001	94	< 0.001	0.328							
2002	72	0.004	0.568	< 0.001						
2003	223	< 0.001	0.244	< 0.001	0.024					
2004	102	< 0.001	0.132	0.006	0.002	< 0.001				
2005	72	0.004	0.808	0.016	0.04	< 0.001	0.002			
2006	120	< 0.001	0.866	0.02	< 0.001	< 0.001	< 0.001	< 0.001		
2007	48	0.004	0.358	0.052	< 0.001	0.002	0.004	< 0.001	0.114	
2008	44	0.666	0.456	0.006	< 0.001	0.004	0.006	0.044	0.022	0.028

TABLE 5.5: P-values of the pairwise comparison of the spatial distribution of natural wildfires between years.

5.3 Analysis of interaction between wildfires

Once confirmed that the spatial distribution of the wildfire patterns are inhomogeneous and estimated their first-order intensity functions, we focus on the second-order characteristics of the wildfire patterns. In this section we analyze the dependence structure within univariate patterns to answer questions such as: "did arson fires registered during 2003 occur independently or were clustered?", as well as the dependence between different types of wildfires, which may allow us to answer questions such as "were small and large fires registered during 2006 clustered?", or "did arson fires in 2006 occur close to those registered in 2005?".

5.3.1 Spatial interaction within wildfire patterns

We have conducted inhomogeneous L-tests (see Section 2.2.6.2) to characterize the dependence structure of the spatial patterns of wildfires (Figure B.21 in Appendix B.2), wildfires by size (Figures B.22 - B.24), and wildfires by cause (Figures B.25 - B.29), for the ten years under study. The inhomogeneous L-function of each observed pattern was compared with the envelopes of $B = 39$ realizations of inhomogeneous spatial Poisson point processes with the same first-order intensity as the observed pattern, to obtain a test with significance level $\alpha = 0.05$.

Table 5.6 shows the type of dependence detected for each observed pattern and the radius of interaction. The inhomogeneous L-test detected clustering in the spatial patterns of wildfires with a radius of interaction of 10 km for the ten years under study (Figure B.21), this implies that the presence of a wildfire in a given location favors the occurrence of new ignitions at distance lower than 10 km. The same result was found for small and regular fires, although the radius of interaction raised up to 15 km, (Figures

B.22 and B.23). In contrast, the L-test did not detect spatial interaction between large fires, expect during 2005 and 2006 that exhibited aggregation up to 15 km (Figure B.24).

The inhomogeneous L-test applied to wildfires classified by cause detected clustered patterns for human-caused fires: arson (Figure B.25) and negligences (Figure B.27), and wildfires with unknown cause (Figure B.29). While our results indicate that natural wildfires occurred independently, except in 2006 and 2008 where positive interaction was found (Figure B.26).

			1999	2000	2001	2002	2003	2004	2005	2006	2007	2008
	Wildfires	Type	A	A	A	A	A	A	A	A	A	A
		(km)	10	10	10	10	10	10	10	10	10	10
Size	$S < 1$	Type	A	A	A	A	A	A	A	A	A	A
		(km)	15	15	15	15	15	15	15	15	15	15
	$1 \leq S < 25$	Type	A	A	A	A	A	A	A	A	A	A
		r (km)	15	15	15	15	15	15	15	15	10	15
	$S \geq 25$	Type	I	I	I	I	I	I	A	A	I	I
		r (km)							15	15		
Cause	arson	Type	A	A	A	A	A	A	A	A	A	A
		(km)	15	15	15	15	15	15	15	10	15	15
	natural	Type	I	I	I	I	I	I	I	A	I	A
		(km)								10		10
	negligence	Type	A	A	A	A	A	A	A	A	A	I
		(km)	25	25	5	20	20	5	5	15	5	
	reproduction	Type	I	A	A	I	A	A	I	A	I	I
		(km)		15	5		15	5		10		
	unknown	Type	A	A	A	A	A	A	A	A	A	A
		r (km)	15	15	10	15	10	5	10	20	15	20

TABLE 5.6: Inhomogeneous L-test for the spatial patterns of wildfires registered in Galicia from 1999 to 2008 classified by size and cause. Type of interaction: A: aggregated pattern, I: independent events, and R: regular pattern.

5.3.2 Spatial interaction between different patterns

Spatial interaction between two types of events occurs when different types of events are either closer or further apart than expected under the assumption that the marginal processes are independent. In this Section we applied Monte Carlo tests based on the inhomogeneous L-cross function (Section 2.3.3) to test the dependence between small, regular and large fires, in order to check whether the occurrence of small fires can be considered a risk factor for large fires. We have also analyzed the relationship of arson

fires with natural fires, negligences and reproductions.

Figures 5.6 and 5.7 show positive interaction between small and regular wildfires in the 10 years considered in this study. On the other hand, we do not observe evidences of interaction between small and large fires, and between regular and large fires during the first half of the study period. While from 2004 to 2006 the Monte Carlo test identified aggregated patterns for both pairs. Notice that the small number of large fires in 1999, 2007 and 2008 lead to high variability in the empirical L-cross function and hampers the interpretation of the graphical test. The positive interaction between small and large fires indicates that the occurrence of a small fire in a given location increases the risk of large fires in its neighborhood. This pattern agrees with the behavior of arsonists that aim to burn an area, and repeatedly ignites fires in the area; these fires are extinguished by the Fire Fighting Service and do not lead to large fires until the service no longer has the capacity to respond and the arsonist finally achieves his/her objective.

The analysis of interaction between arson fires and fires with other causes (Figures 5.8 and 5.9) did not find any dependence between arson and natural fires. But the L-cross tests reported positive interaction between arson fires and those due to negligence, the two types of human-caused fires, and between arson fires and reproductions.

The inhomogeneous L-index (Genton et al., 2006), which enables test presentation for several pairs of patterns in a single plot, was used to analyze the spatial dependence between pairs of sequential years for large fires, arson fires with more than 1 ha burned, and natural fires, and to analyze the relationship between wildfires registered in sequential weeks in a given year. The L-index is an estimate of the area between the L-cross function of the observed pattern, $L_{cross,ij}(r)$, and the reference line, $r = r$. This index is obtained by summing the distances between the empirical L-cross function and the main diagonal up to a given distance, $L_{index,ij} = \sum_{r<r_{max}} (L_{cross,ij}(r) - r)$. In this work $r_{max} = 15$ km was fixed as maximum distance.

The analysis of interaction between year for large fires (Figure 5.10) suggests inhibition between consecutive years from 2000 to 2003 and aggregation from 2003 to 2006. We also found evidences of inhibition between 2001 and 2003, and positive interaction for the pairs 2004-2006 and 1999-2005.

Comparison between the L-indices for arson fires with more than 1 ha burned (Figure 5.11) and natural fires (Figure 5.12) reveal two different behaviors. For arson fires

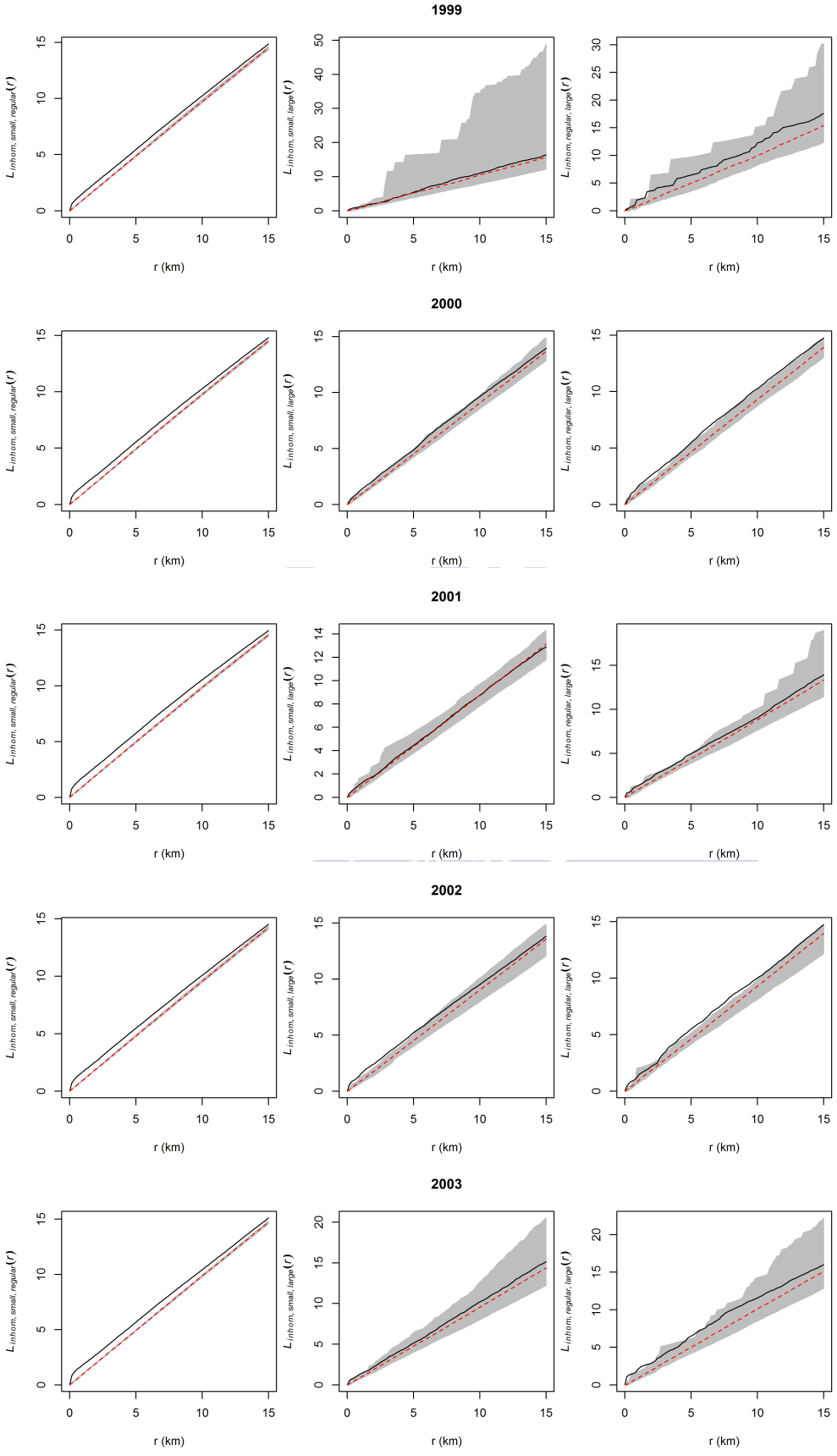


FIGURE 5.6: Inhomogeneous L-cross for pairwise interactions between wildfires by size (1999-2003). Red line: theoretical L-cross: black line: observed L-cross; grey: confidence region delimited by the upper and lower envelopes.

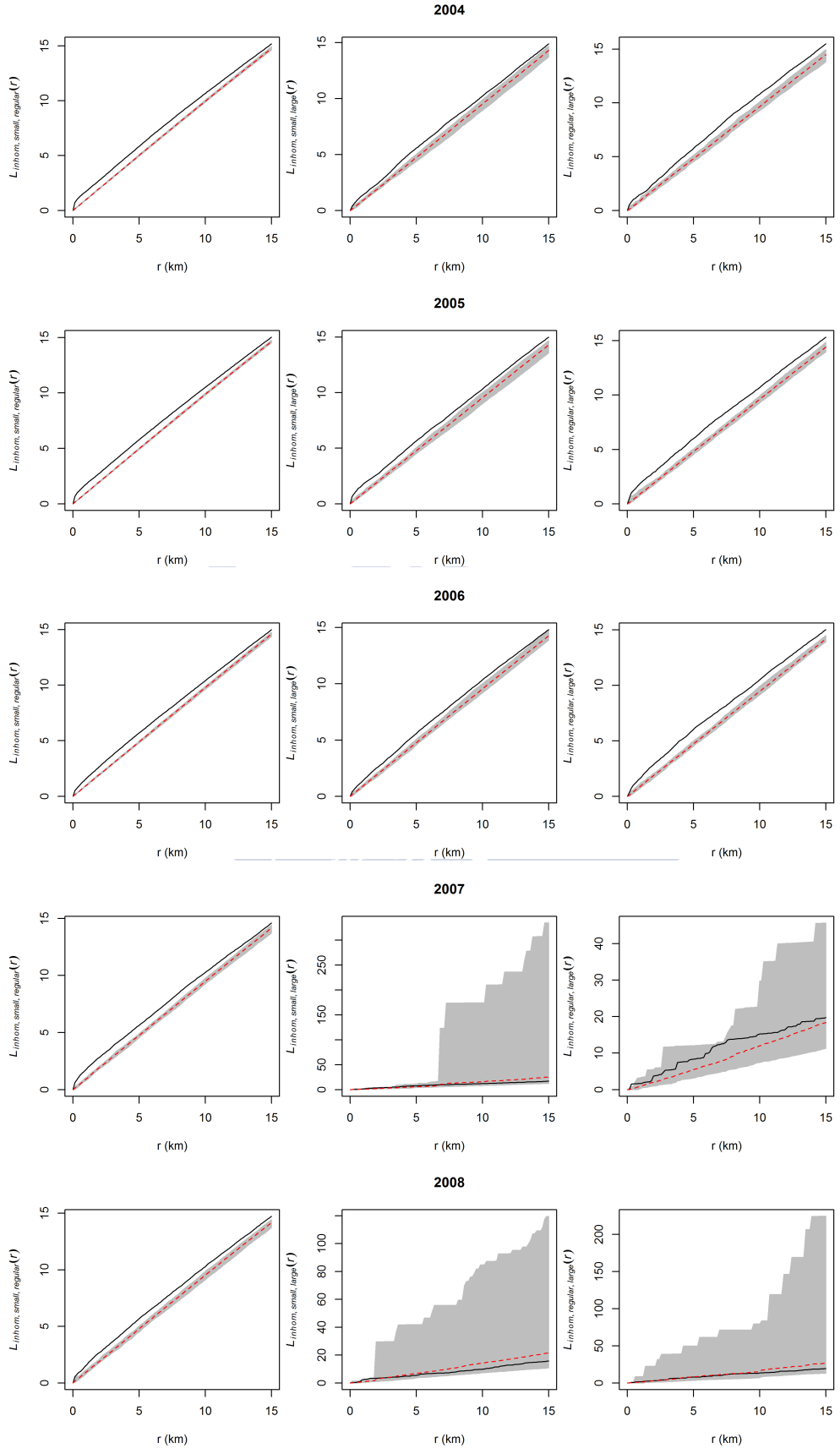


FIGURE 5.7: Inhomogeneous L-cross for pairwise interactions between wildfires by size (2004-2008). Red line: theoretical L-cross: black line: observed L-cross; grey: confidence region delimited by the upper and lower envelopes.

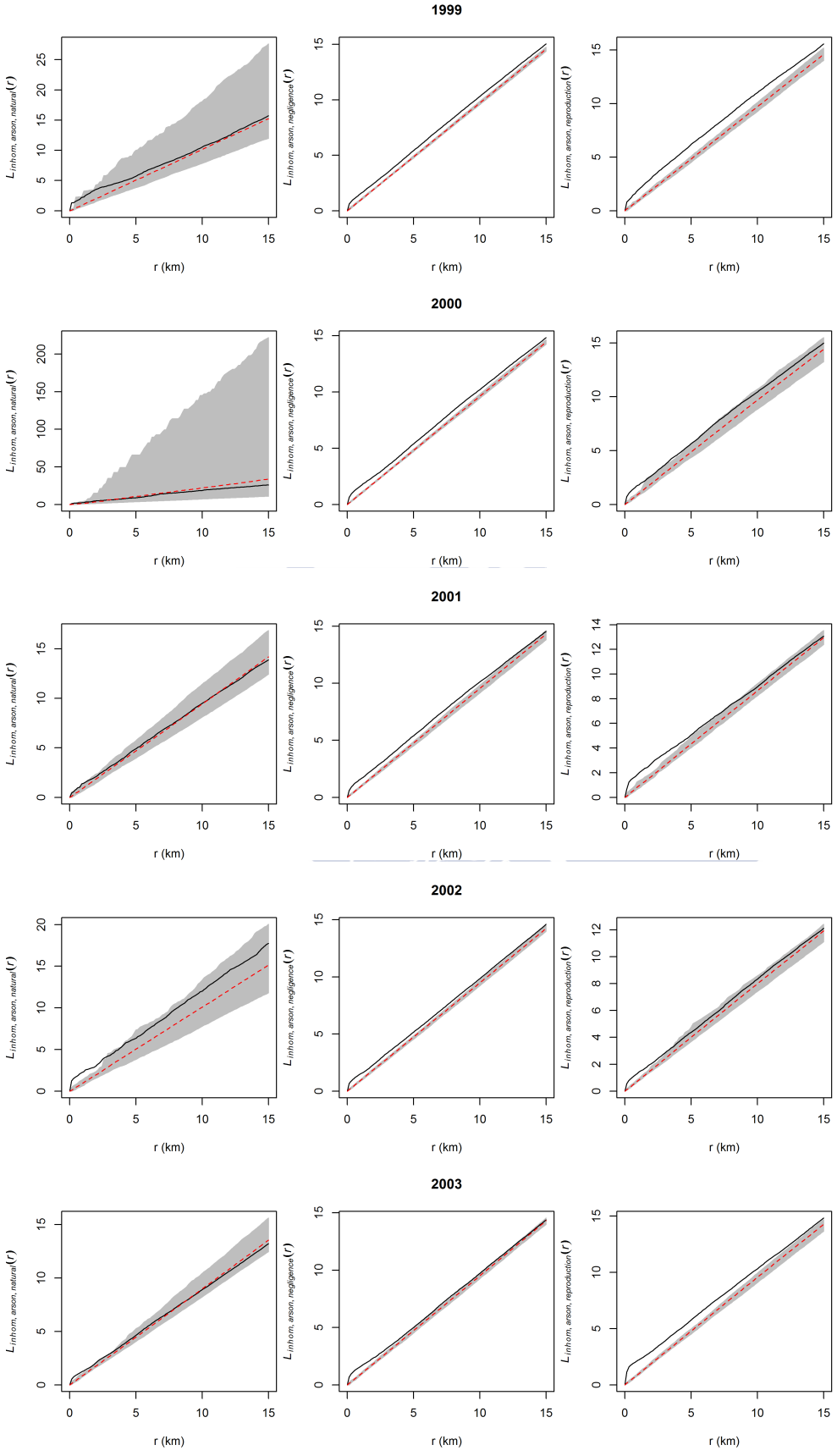


FIGURE 5.8: Inhomogeneous L-cross for pairwise interactions of arson wildfires with natural fires (left),fires cause by negligence (center) and reproductions (right), (1999-2003). Red line: theoretical L-cross: black line: observed L-cross; grey: confidence region delimited by the upper and lower envelopes.

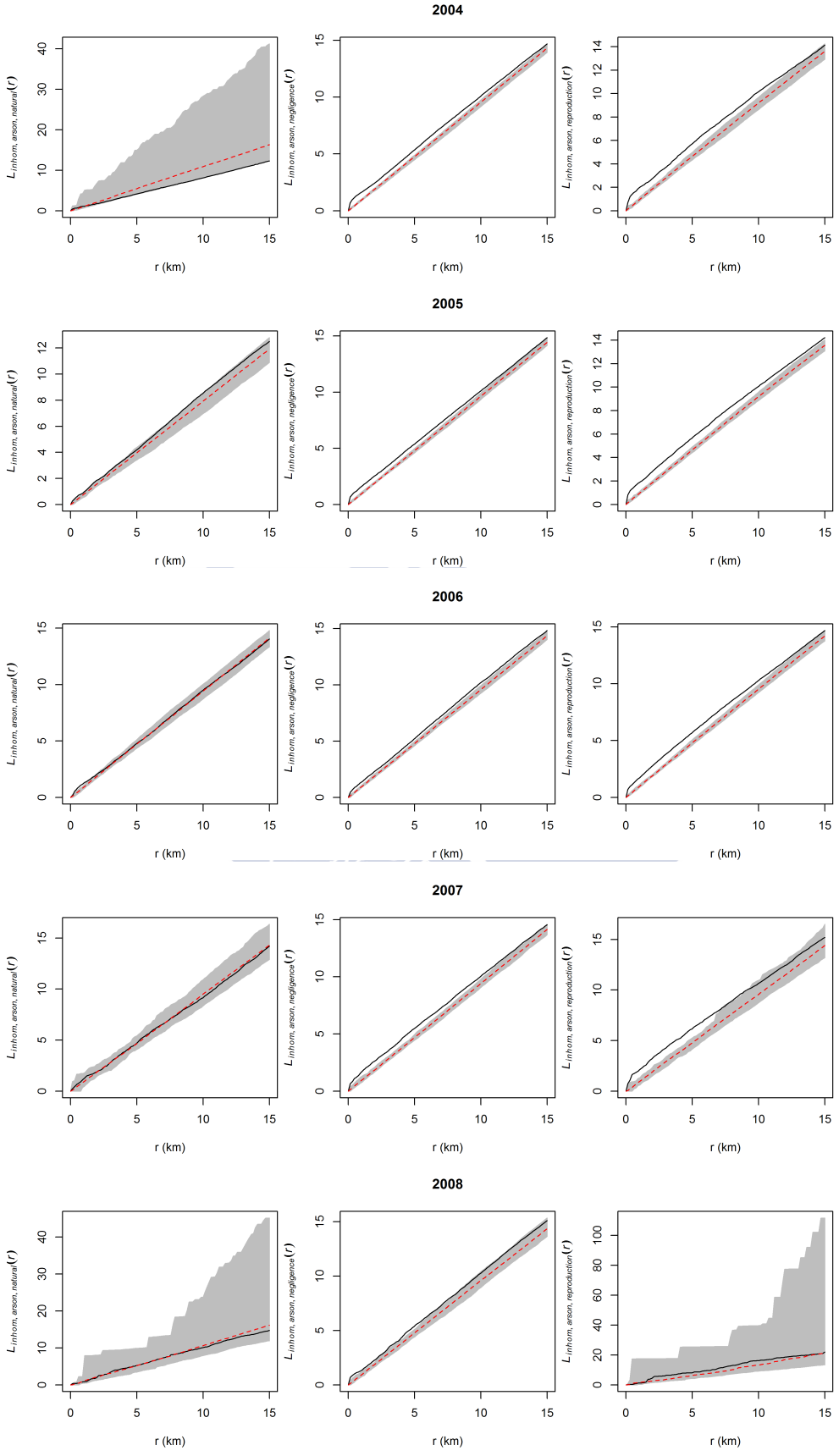


FIGURE 5.9: Inhomogeneous L-cross for pairwise interactions of arson wildfires with natural fires (left), fires caused by negligence (center) and reproductions (right), (2004-2008). Red line: theoretical L-cross: black line: observed L-cross; grey: confidence region delimited by the upper and lower envelopes.

we found evidences of aggregation, except between the pairs 2006-2007 and 2006-2008, which had negative interaction. While the occurrence of natural fires over sequential years tended to be independent, except for the positive interaction found in the pairs 2001-2003, 2003-2008 and 2004-2008, and the inhibition detected between the natural fires registered in 2003 and 2005. Notice that we did not find any interaction between the pairs 2006-2007 and 2007-2008, which according to the nonparametric comparison conducted in Section 5.2.2 had the same spatial distribution.

We have also analyzed the interactions between wildfires registered in sequential weeks during 2004 and 2006. Figures 5.13 and 5.14 show positive interaction between consecutive weeks in both years. Although the evidence of interaction reduces with the temporal lag between fires, Figure 5.13 shows aggregation between wildfires occurring four weeks apart in 2004.

5.4 Conclusion

This chapter illustrates the application of spatial point processes to the analysis of wildfires. The first-order analysis detected spatial variability in the distribution of wildfires in Galicia, and the kernel intensity estimators allowed us to identify the areas with higher incidence of different types of fires. The nonparametric comparison between spatial patterns found that the spatial distribution of large, arson and natural fires changed over years.

The positive interaction detected between small, regular and large fires, agrees with the behavior of arsonist and suggests that small fires can be used to predict the risk of large fires at short-term scale. The lack of interaction between arson and natural fires, and the differences in their spatial distribution (see Section 4.4), suggest that different covariates and models should be used to predict the occurrence of each type of events. These differences between arson and natural wildfires and the high incidence of arson fires in Galicia may explain the poor performance of risk indexes based in environmental conditions, which goodness-of-fit has been proved in other countries, in Galicia.

Finally the differences found in the spatial distribution of wildfires over years and over months (Section 4.4), and the interactions between fires occurring in consecutive years and weeks. suggest that we should take into consideration the time of occurrence of ignition points and use spatio-temporal point processes to analyze the wildfire dataset.

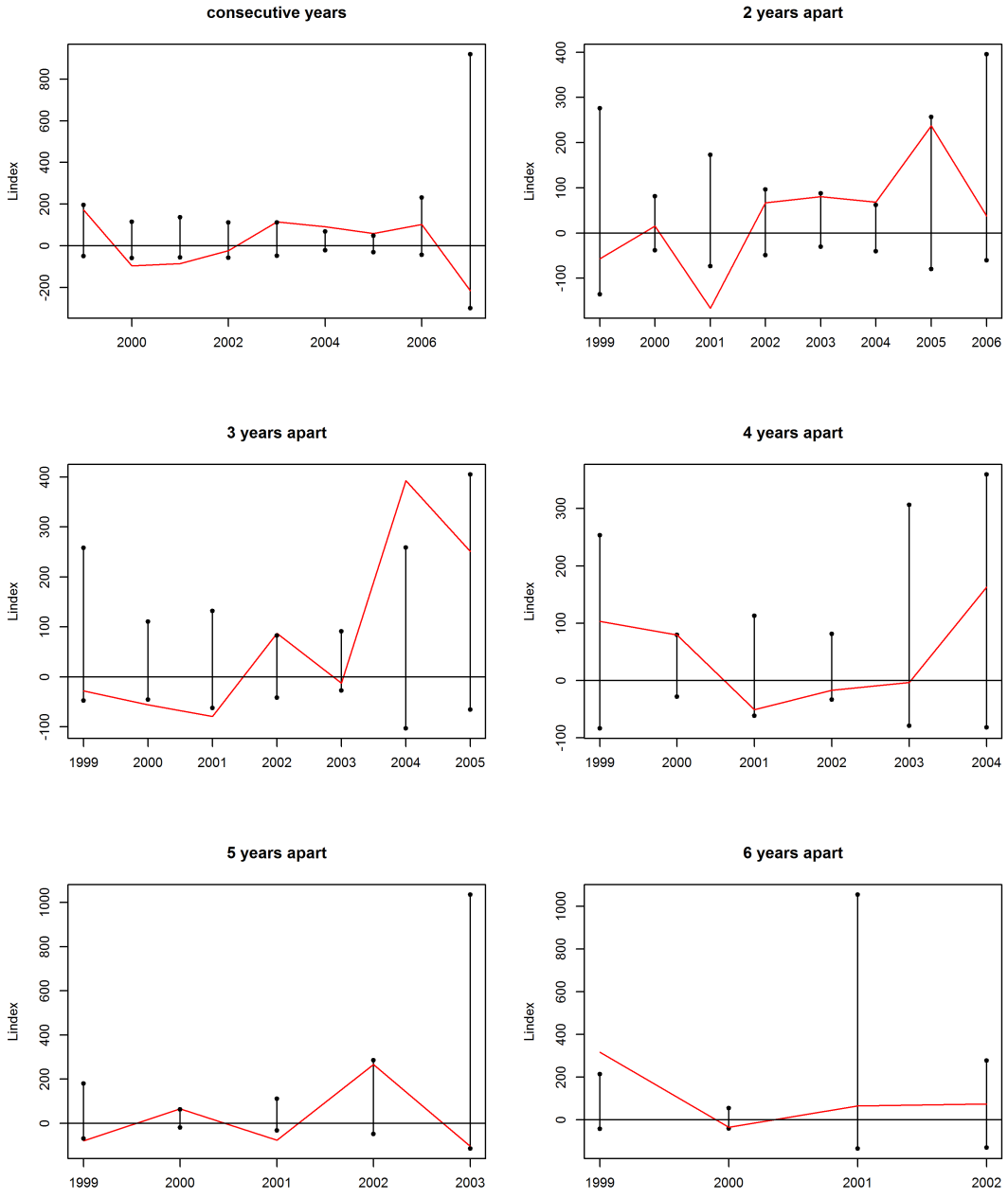


FIGURE 5.10: L-index to test for spatial interactions between large wildfires ($S > 25$ ha over sequential years. Black segments: interval defined by the L-indexes of the upper and lower envelopes, red line: L-index for the observed patterns.

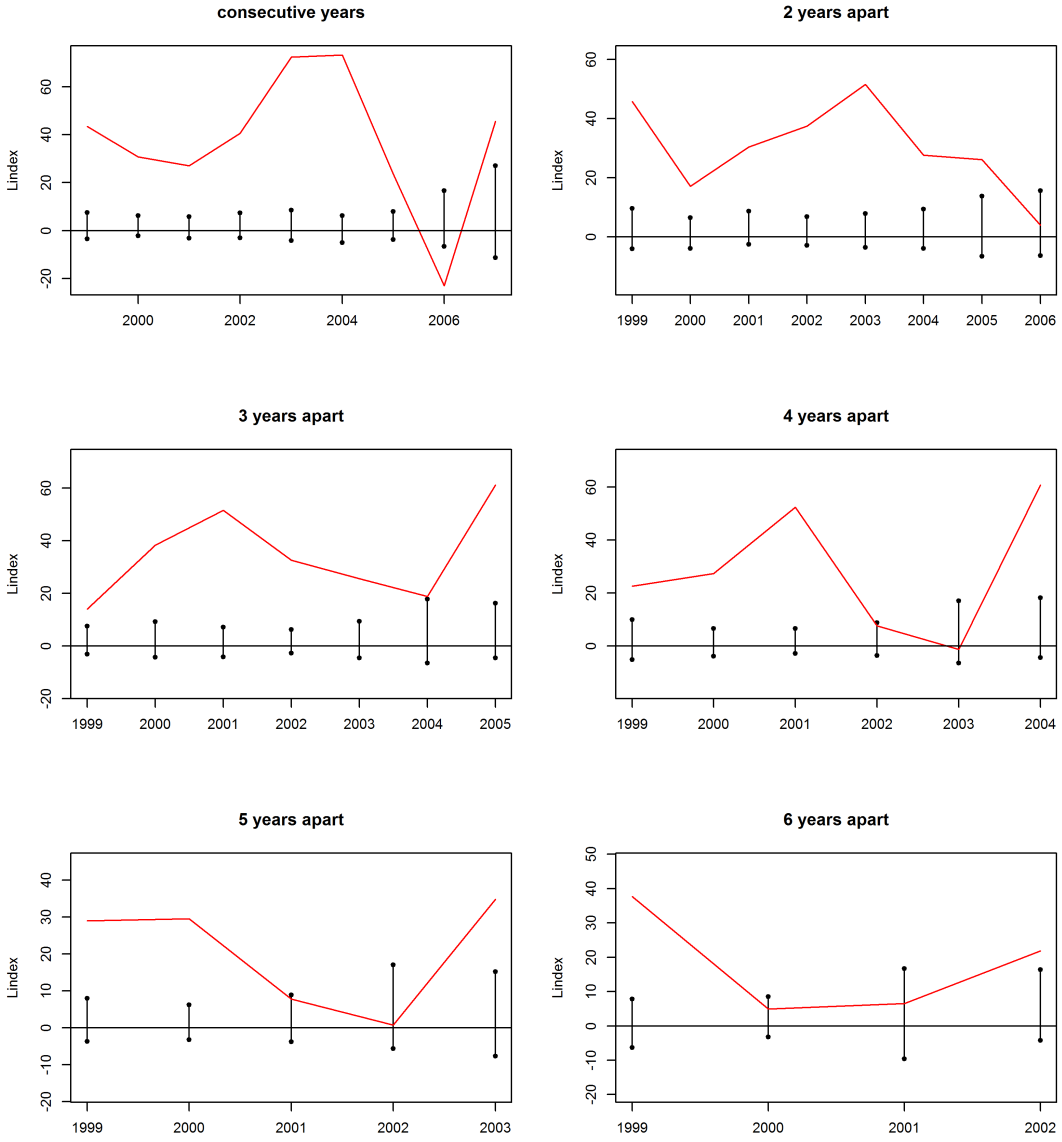


FIGURE 5.11: L-index to test for spatial interactions between arson wildfires over sequential years. Black segments: interval defined by the L-indexes of the upper and lower envelopes, red line: L-index for the observed patterns.

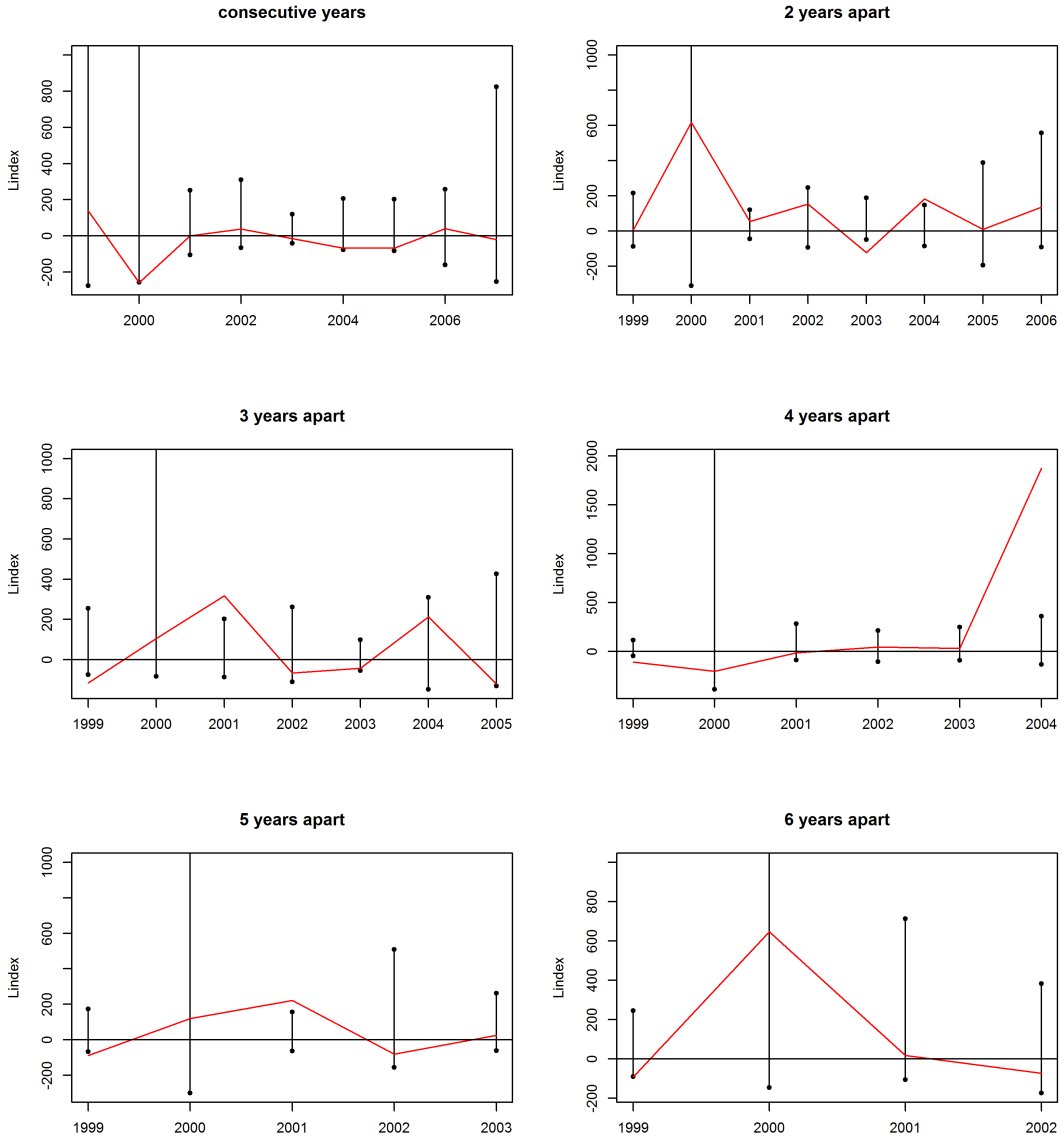


FIGURE 5.12: L-index to test for spatial interactions between natural wildfires over sequential years. Black segments: interval defined by the L-indexes of the upper and lower envelopes, red line: L-index for the observed patterns.

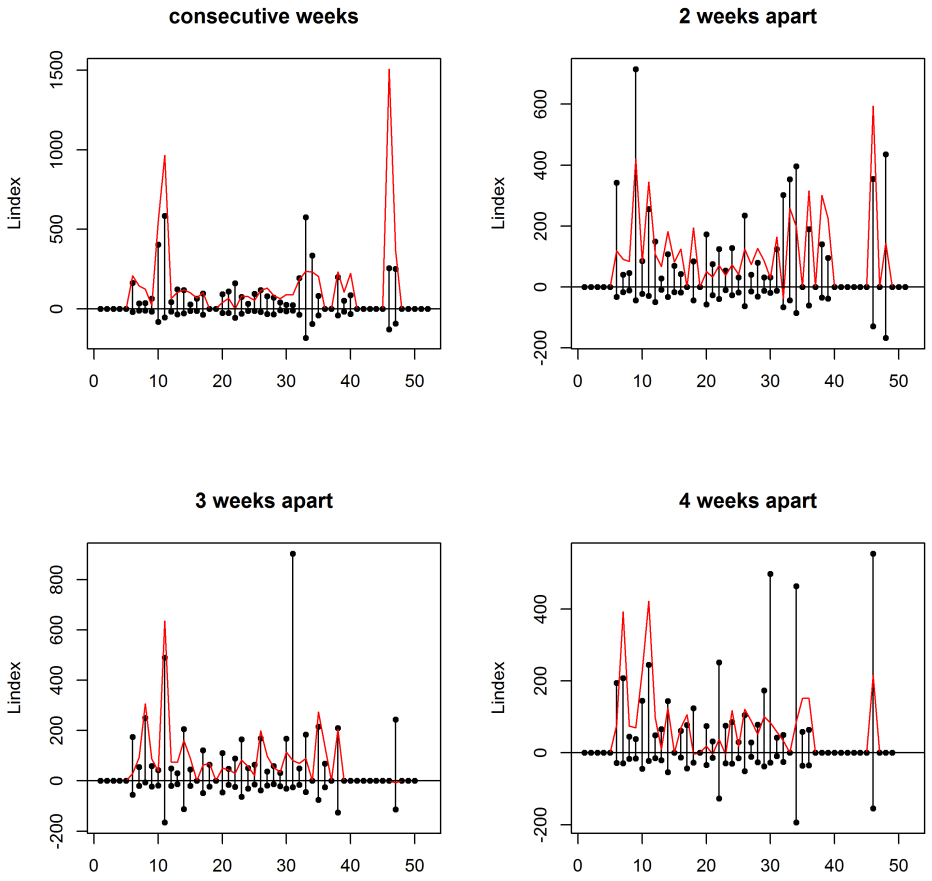


FIGURE 5.13: L-index to test for spatial interactions between the wildfires registered during 2004 over sequential weeks. Black segments: interval defined by the L-indexes of the upper and lower envelopes, red line: L-index for the observed patterns.

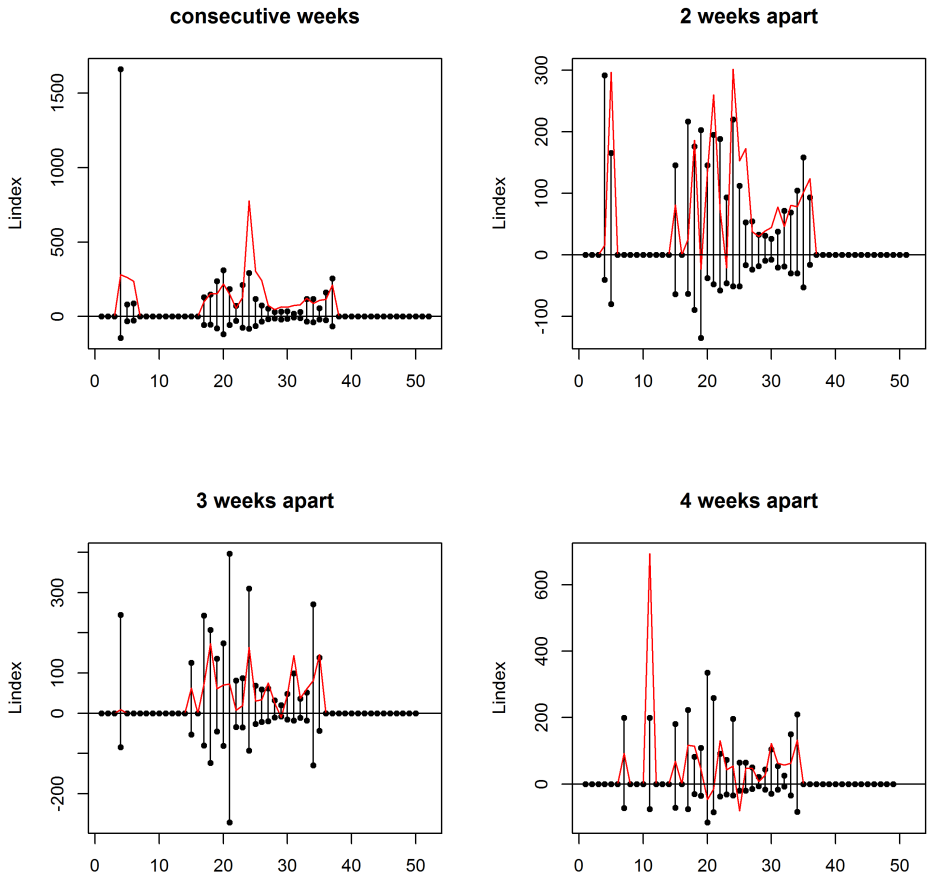


FIGURE 5.14: L-index to test for spatial interactions between the wildfires registered during 2006 over sequential weeks. Black segments: interval defined by the L-indexes of the upper and lower envelopes, red line: L-index for the observed patterns.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Chapter 6

Nonparametric separability test for spatio-temporal point processes

6.1 Introduction

Until now we have focused on the first-order characteristics of spatial point processes. Most spatial point processes in nature are snapshots or cumulated distributions of spatio-temporal point processes, for instance the spatial pattern of wildfires registered a particular day or along a year. Although analyzing this kind of datasets as purely spatial data provides useful in some cases, we should take advantage of the additional information provided by the temporal component using spatio-temporal methods.

Let $\mathbf{S} = \{(\mathbf{x_1}, \mathbf{t_1}), \ldots, (\mathbf{x_N}, \mathbf{t_N})\}$ be a partial realization of a continuous spatio-temporal point process. For instance, the spatio-temporal pattern comprising the spatial locations and time of occurrence of the wildfires observed in a given region, W, during the period T. As in the spatial framework, modeling the spatio-temporal intensity function (2.4.1)

$$\lambda(x,t) = \lim_{|dx \times dt| \to 0} \left\{ \frac{E\left[N(dx,dt)\right]}{|dx \times dt|} \right\}$$

which measures the expected rate of events occurrence per unit space-time volume, and consequently characterizes the first-order structure of any spatio-temporal point process, is one of the main issues in the analysis of any observed pattern. However, modeling the joint distribution of spatial locations and time of occurrence can be a challenging

task, whose complexity increases if the point process is marked or depends on covariates.

Most of the current models assume a separable spatio-temporal intensity function (see details in Section 2.4.2. This hypothesis allows to express the intensity function in a product form $\lambda(x,t) = \lambda_1(x)\lambda_2(t)$, easing its estimation and avoiding the curse of dimensionality. However in the application to real data the separability assumption can be quite restrictive and unrealistic. For instance, a separable model for wildfires would assume that the spatial distribution of wildfire risk does not vary over time. Therefore, testing separability should be one of the first steps in the analysis of any observed spatio-temporal point pattern. Recently some works (Schoenberg, 2004; Assuncáo and Maia, 2007; Chang and Schoenberg, 2011; Díaz-Avalos et al., 2013) developed nonparametric Monte Carlo tests to check whether the spatio-temporal conditional intensity function (2.4.3) is separable, which can be directly adapted to test for separability in the spatio-temporal intensity function (see details in Sections 2.4.2 and 2.4.3).

The aim of this chapter is to develop a nonparametric separability test which does not depend on Monte Carlo simulations of the null hypothesis. To this purpose, we took into account that from equation (2.4.8), the ratio $r(x,t) = \lambda(t,x)/\lambda_1(x)$ is spatially invariant for any separable point process. Thus, we can estimate this ratio and check whether it depends on the spatial locations through a no-effect test. In Section 6.2 we propose a kernel estimator of $r(x,t)$, analyze its asymptotic properties and propose a bandwidth selector. In Section 6.3 we introduce the nonparametric separability test. Section 6.4 discusses the results of the simulation studies conducted to analyze the performance of our separability test. In Section 6.5 we present the practical application to the spatio-temporal pattern of wildfires registered in Galicia during 2006. The chapter ends with some conclusions in Section 6.6

6.2 Kernel estimation of the log-ratio function

Let $\mathbf{S} = \{(\mathbf{x_i}, \mathbf{t_i}), i = 1, \ldots, N\} \in W \times T \subset \mathbb{R}^2 \times \mathbb{R}^+$ be a realization of an inhomogeneous spatio-temporal Poisson point process, and $X = \{\mathbf{x_i}, i = 1, \ldots, n\} \in \mathbb{R}^2$ the corresponding marginal spatial point pattern. Conditional on the number of events $N = n$, $\mathbf{S}$ and $\mathbf{X}$ can be seen as random samples of the distributions with probability density functions $f(x,t) = \lambda(x,t)/m$ and $g(x) = \lambda_1(x)/m$, where $m = \int_T \int_W \lambda(x,t)dxdt = \int_W \lambda_1(x)dx$ is the expected number of events of both the spatial and spatio-temporal point processes.

The ratio

$$r(x,t) = \frac{\lambda(x,t)}{\lambda_1(x)} = \frac{f(x,t)}{g(x)}$$

can be seen as a spatio-temporal relative risk function which control distribution remains constant over time.

In the same line as the kernel estimators of the spatial and spatio-temporal relative risk functions (see details in Sections 2.3.4 and 2.4.4) we can apply kernel smoothing to estimate the log-ratio function $\rho(x,t) = log(f(x,t)/g(x))$.

The kernel estimator of the spatio-temporal density function is

$$\hat{f}_{H_s,h_t}(x,t) = \frac{\sum_{i=1}^{n} k_s\left(H_s^{-1/2}\left(x - \mathbf{x_i}\right)\right) k_t\left(h_t^{-1}\left(t - \mathbf{t_i}\right)\right)}{n p_{H_s,h_t}(x,t)|H_s|^{1/2} h_t} \tag{6.2.1}$$

where the kernel functions, $k_s(\cdot)$ and $k_t(\cdot)$, are defined using a spherically symmetric bivariate density function and a univariate density function, which in this case is also assumed to be symmetric, and $\int_{\mathbb{R}^2} uu^T k_s(u)du = \mu_2(k_s)I_2$ and $\int_{\mathbb{R}} v^2 k_t(v)dv = \mu_2(k_t)$, with $\mu_2(k_s), \mu_2(k_t) < \infty$. H_s is the two-dimensional bandwidth matrix for the spatial component, and h_t is the unidimensional bandwidth parameter for the temporal component; $p_{H_s,h_t}(x,t) = \int_T \int_W k_{s,H_s}(x-u)k_{t,h_t}(t-v)dxdt = \int_W k_{s,H_s}(x-u)dx \int_T k_{t,h_t}(t-v)dt$ is the spatio-temporal edge-correction term, where $p_{H_s}(x) = \int_W k_{s,H_s}(x-u)dx$ and $p_{h_t}(t) = \int_T k_{t,h_t}(t-v)dt$ represent, respectively, the bivariate edge-corrector for the spatial locations and the univariate edge-corrector for the temporal component.

The kernel estimator of the spatial density function is

$$\hat{g}_H(x) = \frac{\sum_{i=1}^{N} k_s\left(H^{-1/2}\left(x - \mathbf{x_i}\right)\right)}{n p_H(x)|H|^{1/2}} \tag{6.2.2}$$

where, $k_s(\cdot)$ is the same bivariate kernel used in the spatial component of $\hat{f}_{H_s,h_t}(x,t)$, H is a bivariate bandwidth matrix. $p_H(x) = \int_W k_{s,H}(x-u)dx$ is the edge-correction term.

Expression (6.2.1) and (6.2.2) allow us to define the following estimator for the log-ratio function, $\rho(x,t) = \log\left(f(x,t)/g(x)\right)$:

$$\hat{\rho}(x,t) = log\left(\frac{\hat{f}_{H_s,h_t}(x,t)}{\hat{g}_H(x)}\right) = \log \hat{f}_{H_s,h_t}(x,t) - \log \hat{g}_H(x) \tag{6.2.3}$$

There are three issues that deserve attention in the log-ratio kernel estimator: data sparseness, bandwidth selection, and edge-effects. The log-ratio estimator is highly unstable or even undefined in areas where the spatio-temporal density is close to 0, i.e. in regions with sparse data. In these situations, we may need large bandwidths to control the variance of the kernel estimates in areas with low intensity, resulting in excessive blurring of detail elsewhere. This problem has led some authors (Bithell, 1990; Hazelton and Davies, 2009) to suggest the addition of a stabilizing constant, δ, in both the numerator and denominator of $\hat{\rho}(x,t)$, which yields

$$\hat{\rho}(x,t) = \log\left(\frac{\hat{f}_{h_s,h_t}(x,t)+\delta}{\hat{g}_h(x)+\delta}\right) \tag{6.2.4}$$

The bandwidth parameter is crucial for the performance of any kernel estimator. Kernel smoothing of spatial and spatio-temporal intensity and relative risk functions, has been limited to the use of scalar bandwidths for the spatial component, which can be very restrictive in practice specially for highly inhomogeneous and anisotropic patterns. In addition, to obtain the kernel log-ratio function we can use different spatial bandwidths in the numerator and denominator of $\hat{\rho}(x,t)$. However Kelsall and Diggle (1995a), for the spatial relative risk, and Sarojinie Fernando and Hazelton (2014) for the spatio-temporal relative risk, suggested using a common bandwidth for cases and controls, as it leads to a bias cancellation in regions where $f = g$ and simplifies data-driven bandwidth selection. Taking into account these considerations, in this chapter we use diagonal bandwidth matrices to smooth the spatial component, which are less restrictive than single bandwidths and easier to estimate than full bandwidth matrices, and consider the same spatial bandwidth in the kernel estimators of $f(x,t)$ and $g(x)$, $h_s = (h_{s1}, h_{s2}) = dgH_s^{1/2}$. Thus, expression (6.2.3) can be rewritten as follows

$$\hat{\rho}(x,t) = log\left(\frac{\hat{f}_{h_s,h_t}(x,t)+\delta}{\hat{g}_{h_s}(x)+\delta}\right) = \log\left(\hat{f}_{h_s,h_t}(x,t)+\delta\right) - \log\left(\hat{g}_{h_s}(x)+\delta\right) \tag{6.2.5}$$

Another challenging issue is the boundary effect, which arises when data are observed in bounded domains. We have seen above that the spatio-temporal edge-corrector can

be decomposed into its spatial, $p_{h_s}(x)$, and temporal, $p_{h_t}(t)$, components. In some situations, for instance when we observe all cases in some specific epidemic, we know in advance that there are no events outside the time interval T and temporal edge-correction is not necessary. When we observe the pattern over T but there may be events outside the interval, the edge-correction in time is required. In the first case, $p_{h_s,h_t}(x,t) = p_{h_s}(x)$, and the edge-correction terms of the kernel estimators (6.2.1) and (3.2.1) cancel out in expression (6.2.3), i.e. we do not need to compute the edge-correctors in $\hat{f}_{h_s,h_t}(x,t)$ and $\hat{g}_{h_s}(x)$ to estimate the log-ratio function.

6.2.1 Asymptotic properties of the kernel log-ratio function

The performance of $\hat{\rho}(x,t)$ can be measured in terms of its mean integrated square error, defined as follows:

$$\begin{aligned} MISE(h_s,h_t) &= E\left[\int_T\int_W \left(\hat{\rho}(x,t)-\rho(x,t)\right)^2 dxdt\right] \\ &= \int_T\int_W B\left(h_s,h_t\right)^2 dxdt + \int_T\int_W Var\left(h_s,h_t\right) dxdt \end{aligned}$$

where $B(h_s,h_t)$, $Var(h_s,h_t)$ are the bias and variance of $\hat{\rho}(x,t)$ and, as we have conditioned to the number of events, $N=n$, E denotes expectation over the randomness in the spatial location and time of occurrence of events.

To obtain the asymptotic expression of $MISE(h_s,h_t)$ we assume the infill asymptotic framework (see discussion in Section 3.2), which dictates that the expected number of events $m = \int_T\int_W \lambda(x,t)dxdt = \int_W \lambda_1(x)dx$ tends to ∞ as h_s and h_t tend to 0. Therefore we assume that $n \to \infty$ as the bandwidth parameters tend to 0, which is the typical asymptotic framework in kernel density estimation.

Assuming that: (i) f and g are bounded away from 0 and have bounded and square integrable partial derivatives up to order 2, (ii) $k_s(\cdot)$ and $k_t(\cdot)$ are continuous, symmetric, square integrable density functions, then when $n \to \infty$, h_s, $h_t \to 0$ such that $nh_s^2, nh_t \to \infty$, the bias of $\hat{f}_{h_s,h_t}(x,t)$ in any interior point of $W \times T$ is

$$B\left[\hat{f}_{h_s,h_t}(x,t)\right] = \frac{1}{2}\mu_2(k_s)h_s^2\nabla_x^2 f(x,t) + \frac{1}{2}\mu_2(k_t)h_t^2\nabla_t^2 f(x,t) + o\left(h_{s1}^2 + h_{s2}^2 + h_t^2\right)$$

where $\nabla_x^2 f(x,t)$ denotes the column vector $\left(\frac{\partial^2 f(x,t)}{\partial x_1^2}, \frac{\partial^2 f(x,t)}{\partial x_2^2}\right)^t$, and $h_s = (h_{s1}, h_{s2})$. The variance of $\hat{f}_{h_s,h_t}(x,t)$ in any interior point of $W \times T$ is

$$Var\left[\hat{f}_{h_s,h_t}(x,t)\right] = \frac{f(x,t)R(k_s)R(k_t)}{nh_{s1}h_{s2}h_t} + o\left((nh_{s1}h_{s2}h_t)^{-1}\right)$$

The bias and variance of the kernel density estimator for the spatial component are

$$B\left[g_{h_s}(x)\right] = \frac{1}{2}\mu_2(k_s)h_s^2\nabla^2 g(x) + o(h_{s1}^2 + h_{s2}^2)$$

and

$$Var\left[g_{h_s}(x)\right] = \frac{g(x)R(k_s)}{nh_{s1}h_{s2}} + o\left((nh_{s1}h_{s2})^{-1}\right)$$

Therefore, the bias and variance of $\hat{\rho}(x,t)$ are (see details in Appendix C.1):

$$B\left(h_s,h_t\right) = \frac{1}{2}\left[\frac{\mu_2(k_s)h_s^2\nabla_x^2 f(x,t) + h_t^2\mu_2(k_t)\nabla_t^2 f(x,t)}{f(x,t)} - \frac{\mu_2(k_s)h_s^2\nabla^2 g(x)}{g(x)}\right] + o\left(h_{s1}^2 + h_{s2}^2 + h_t^2\right) \tag{6.2.6}$$

$$Var\left(h_s,h_t\right) = \frac{R(k_s)R(k_t)f(x,t)^{-1}}{n_1 h_s^2 h_t} + \frac{R(k_s)g(x)^{-1}}{n_2 h_s^2} + o\left(\left(nh_{s1}h_{s2}\left(1+h_t\right)\right)^{-1}\right) \tag{6.2.7}$$

When time is recorded at a relatively coarse level, the temporal bandwidth, h_t, will be larger than the spatial bandwidth, h_s and the second term in the bias, (6.2.6), will be the dominant term.

From expressions (6.2.6) and (6.2.7) we have that the MISE of $\hat{\rho}(x,t)$ is

$$MISE\left(h_s,h_t\right) = AMISE\left(h_s,h_t\right) + o\left(\left(nh_{s1}h_{s2}\left(1+h_t\right)^{-1}\right) + \left(h_{s1}^2 + h_{s2}^2 + h_t^2\right)^2\right) \tag{6.2.8}$$

where

$$AMISE\left(h_s,h_t\right) = \frac{1}{4}\int_T\int_W\left[\frac{\mu_2(k_s)h_s^2\nabla_x^2 f(x,t) + \mu_2(k_t)h_t^2\nabla_t^2 f(x,t)}{f(x,t)} - \frac{h_s^2\mu_2(k_s)\nabla_x^2 g(x)}{g(x)}\right]^2 dxdt + \int_T\int_W\left[\frac{R(k_s)R(k_t)f(x,t)^{-1}}{n_1 h_{s_1}h_{s_2}h_t} + \frac{R(k_s)g(x)^{-1}}{n_2 h_{s_1}h_{s_2}}\right]dxdt \tag{6.2.9}$$

The AMISE could be used as a discrepancy measure to develop data-driven bandwidth selection procedures. Given the good reputation of plug-in bandwidth selectors in kernel density estimation (Duong and Hazelton, 2003; Chacón and Duong, 2011) and in kernel intensity estimation (see Chapter 3), we can consider developing a plug-in bandwidth selector based on minimizing expression (6.2.9) for the kernel log-ratio estimator. However, the need of robust estimates for reciprocals of density functions hampers this idea.

6.2.2 Bandwidth selection

In addition to the well-known effect of the bandwidth parameters on the goodness-of-fit of any kernel estimator, the degree of smoothness of the kernel log-ratio function plays an important role on the power of the separability test. Small bandwidths may lead to identify slight changes in the log-ratio function of a separable point process as non-separability, while too large bandwidths may not detect departure from separability.

Previous nonparametric tests (Schoenberg, 2004; Díaz-Avalos et al., 2013) made a subjective selection of the bandwidth parameters in the kernel estimators of the separable and nonseparable spatio-temporal intensity functions. In this work, following the methodology applied for kernel relative risk functions (Kelsall and Diggle, 1995a; Sarojinie Fernando and Hazelton, 2014), we propose a least-squares cross-validation (LSCV) bandwidth selector for the log-ratio function. Sarojinie Fernando and Hazelton (2014) argued that the LSCV bandwidth selector (2.4.24) cannot be extended to the context of spatio-temporal log-relative risk functions with time-invariant control distribution (2.4.26) because $f(x,t)$ cannot be evaluated at control data points as controls are not indexed by time. In the log-ratio function the spatial pattern of controls, X, is the spatial marginal of the spatio-temporal pattern of cases S, thus we shall not find this problem.

Let $\gamma(f,g)$ be a smooth functional of the spatio-temporal and spatial densities. A common procedure to select the optimal bandwidth to estimate γ is to minimize its integrated square error (ISE):

$$ISE(h_s,h_t) = \int_T \int_W \left(\gamma\left(\hat{f}_{h_s,h_t}(x,t),\hat{g}_{h_s}(x)\right) - \gamma(f(x,t),g(x))\right)^2 dxdt$$

which is equivalent to minimize:

$$LSCV\,(h_s,h_t) = \int_T\int_W \gamma(\hat{f}_{h_s,h_t}(x,t),\hat{g}_{h_s}(x))^2 dxdt - 2\int_T\int_W \gamma(\hat{f}_{h_s,h_t}(x,t),\hat{g}_{h_s}(x))\gamma(f(x,t),g(x))dxdt \tag{6.2.10}$$

Thus for the log-ratio function $\gamma(f,g) = \log(f/g)$, we obtain the following selection criterion (see details in Appendix C.2).

$$LSCV(h_s,h_t) = -\int_T\int_W \hat{\rho}_{h_s,h_t}(x,t)^2 dxdt - 2n^{-1}\sum_{i=1}^{n}\log\left(\frac{\hat{f}^{-i}_{h_s,h_t}(x_i,t_i)}{\hat{g}_{h_s}(x_i)}\right)\left(\hat{f}^{-i}_{h_s,h_t}(x_i,t_i)\right)^{-1} + 2n^{-1}\sum_{j=1}^{n}\log\left(\frac{\hat{f}_{h_s,h_t}(x_j,t_j)}{\hat{g}^{-j}_{h_s}(x_j)}\right)\left(\hat{g}^{-j}_{h_s}(x_j)\right)^{-1} \tag{6.2.11}$$

where $\hat{f}^{-i}_{h_s,h_t}(\cdot)$ and $\hat{g}^{-j}_{h_s}(\cdot)$ denote the leave-one-out estimators of $f(\cdot)$ and $g(\cdot)$, respectively.

6.3 The separability test

For a separable spatio-temporal point process, $\lambda(x,t) = \lambda_1(x)\lambda_2(t)$ and, consequently the log-ratio function, $\rho(x,t) = log\,(\lambda(x,t)/\lambda_1(x))$ is constant for any $t \in T$. This suggests that testing whether the log-ratio function, $\rho(x,t)$, depends on the spatial location of events would be equivalent to test whether the point process is separable.

Following Bowman and Azzalini (1997) we can use nonparametric regression to fit the log-ratio function evaluated at each event, $Y = \{y_i = \rho(\mathbf{x_i},\mathbf{t_i}); i = 1,\ldots,n\}$, according to the spatial locations of events, $X = \{x_i = (x_{i1},x_{i2}), i = 1,\ldots,n\}$. In order to test for the effect of X on Y we consider two competing models

$$\mathcal{H}_0 : E(y_i) = \mu$$
$$\mathcal{H}_1 : E(y_i) = m(x_i)$$

where $m(\cdot)$ is an unknown smooth function, which can be estimated by the kernel approach introduced by Nadaraya (1964) and Watson (1964)

$$\hat{m}(x_1, x_2) = \frac{\sum_{i=1}^{n} w_{g_1}(x_{i1} - x_1) w_{g_2}(x_{i2} - x_2) y_i}{\sum_{i=1}^{n} w_{g_1}(x_{i1} - x_1) w_{g_2}(x_{i2} - x_2)} \tag{6.3.1}$$

where the kernel, w, is a univariate symmetric density function and $g = (g_1, g_2)$ is the vector of smoothing parameters. Three alternative procedures have been commonly used to select this parameter: (i) bandwidth selector associated to the approximate degrees of freedom, df, of the regression errors, (ii) least-squares cross-validation, and (iii) an AICC-based method (Hurvich et al., 1998).

Once computed $\hat{y} = \sum_{i=1}^{n} y_i$, which is the empirical estimator of μ in $\mathcal{H}_0$, and the regression function, $\hat{m}(\cdot)$ in $\mathcal{H}_1$, we compute the residual sum of squares for the null and alternative models

$$RSS_0 = \sum_{i=1}^{n} (y_i - \bar{y})^2$$

$$RSS_1 = \sum_{i=1}^{n} (y_i - \hat{m}(x_i))^2$$

and define the test statistic:

$$F = \frac{(RSS_0 - RSS_1) / (df_1 - df_0)}{RSS_1 / df_1} \tag{6.3.2}$$

where df_0 and df_1 denote the degrees of freedom for the error under each hypothesis. In the linear model framework, the squared sums of residuals and the test statistic follow, respectively, a χ^2 and a F distribution. However, these properties are not fulfilled in the context of nonparametric regression. Bowman and Azzalini (1997) proposed two procedures to estimate the distribution of F under the null hypothesis: (i) if the errors have normal distribution, the null distribution of F is approximated by a shifted and scaled χ^2; (ii) when the errors are not normal, the distribution of F is approximated throughout a computationally intensive procedure based on permutation tests.

The permutation test relies on the fact that under $\mathcal{H}_0$ the pairing of any particular x and y is completely random. Then, the distribution of the test statistic, F, can be generated by simulation, using random pairings of the observed values of X and Y and computing the corresponding test statistic in each case. The empirical p-value of the test is the proportion of simulated F-statistics larger than that obtained from the observed data. In advance this separability test shall be refereed as *F-test*.

6.4 Simulation study

We have conducted two simulation studies to analyze the performance of the nonparametric separability test introduced above. Our first aim was to check whether the χ^2 approximation, and the permutation test provide accurate calibrations of the test statistic. We also tested the effects of the bandwidth parameters used to estimate $\rho(x,t)$ and $m(x)$ on the F-test. Once determined the best bandwidth selectors for both steps, we conducted a new simulation study to compare the performance of our separability test with the nonparametric tests proposed by Schoenberg (2004) and Díaz-Avalos et al. (2013).

We generated realizations of inhomogeneous spatio-temporal Poisson point processes with $m = 500$, considering two different gradual departures from separability. We first simulated point patterns with intensity function

$$\lambda_1(x,t) = 1000(1-\epsilon)\phi_{2,(\mu,\Sigma)}(x)e^{-t} + 450\epsilon\left(\frac{1}{4}\right)^3 I^3_{[0.5,0.75]}(x,t) \tag{6.4.1}$$

where $\phi_{2,(\mu,\Sigma)}$ is the bivariate normal density with mean $\mu = (0.5, 0.5)$ and covariance matrix $\Sigma = 0.05I_2$, where I_2 is the two-dimensional identity matrix. The degree of departure from separability is determined by ϵ, which ranges from 0 to 0.5 at 0.05 intervals. If $\epsilon = 0$ we simulate realizations of a separable spatio-temporal point process with first-order intensity $\lambda^S(x,t) = 1000\phi_{2,(\mu,\Sigma)}(x)e^{-t}$, and as ϵ increases, a higher proportion of events uniformly distributed in the cube $[0.05, 0.75]^3$ is added to the separable point process with intensity $(1-\epsilon)\lambda^S(x,t)$ (see Figures C.1 - C.3 in Appendix C.3). These point processes shall be refereed as STPP1.

In order to generate a different way of departure from separability, we simulated spatio-temporal point patterns with a spatial distribution shrinking throughout time. For this purpose we first simulated separable point patterns with intensity function

$$\lambda^S_2(x,t) = 1000\phi_{2,(\mu,\Sigma)}(x)e^{-t} \tag{6.4.2}$$

with $\mu = (0.5, 0.5)$ and covariance matrix $\Sigma = 0.1I_2$. Once the separable pattern was simulated, we defined the new spatial coordinates as $(x_{i1}, x_{i2}) = (x^0_{i1}, x^0_{i2})(1-\epsilon t^*_i)$, where $\left\{(x^0_{i1}, x^0_{i2}); i = 1, ..., n\right\}$ is the spatial marginal of the simulated pattern and t^* denotes its temporal component rounded up to the second decimal. As above ϵ ranges from 0 to 0.5 at 0.05 intervals and determines the degree of departure from separability (see

Figures C.4 - C.6). These point processes shall be refereed as STPP2.

The simulation study was conducted with the R statistical software (R Core Team, 2014). To simulate the spatio-temporal point patterns we used the *stpp* package of R (Gabriel et al., 2013). To compute the kernel estimator of the log-ratio function (6.2.4) we extended to the spatio-temporal framework the *risk* function of the *sparr* package (Davies et al., 2011), with $\delta = 10^{-6}$. The kernel estimator (6.2.1) in the numerator of (6.2.4) was computed extending the *bivariate.density* function with fixed bandwidths in the *sparr* package to the spatio-temporal framework using Gaussian kernels in the spatial and temporal components, a diagonal bandwidth matrix in the spatial component, and edge-correction in both components. The kernel density estimator (6.2.2) in the denominator of (6.2.4) was computed using the *density* function in *spatstat* (Baddeley and Turner, 2005). To implement the LSCV bandwidth selector we extended the *LSCVrisk* function in *sparr* using a Newton-Raphson algorithm to obtain the bandwidth vector, $h = (h_{s1}, h_{s2}, h_t)$, minimizing (6.2.11). The no-effect test was conducted using the *sm* package Bowman and Azzalini (2014).

For the different scenarios defined throughout this section, we estimated the test statistic, F, and the corresponding empirical p-value for 200 realizations of each point process. The probability of rejecting separability at any significance level, α, was computed as the proportion of p-values smaller than α.

6.4.1 Comparison of calibration methods

The first simulation study was conducted to test whether the χ^2 approximation provides an accurate calibration of the test statistic, or we need the more intensive permutation test. For this purpose, we compared the performance of the χ^2 approximation, implemented in the *sm.regression*, with the calibration provided by the permutation test, with $B = 500$ realizations of the random pairings.

In this simulation study the kernel log-ratio function (6.2.4) was obtained using the LSCV bandwidth selector introduced above. The kernel regression function (6.3.1) was obtained using the three bandwidth selectors available in the *sm* package, and the spatial component of the LSCV bandwidth, h_s. The effect of the bandwidth parameters on the performance of the separability test are analyzed in Section 6.4.2.

For STPP1 and $\epsilon = 0$ the two calibration methods provided probabilities of rejecting the null hypothesis close to the nominal significance level, although the type 1 errors obtained with the χ^2 calibration where slightly larger than the nominal level except when we use h_s to estimate the regression function 6.1. We can also see that the power of the test increased with the departure from separability with both calibrations (see also Figure C.7). For STPP2 the type I errors obtained with the permutation test are closer to the nominal level than those provided by the χ^2 calibration, which was slightly anti-conservative. The power of the test increased with the departure from separability with both calibrations, but the permutation test was more powerful (see Figure C.7).

In view of these results, hereafter we shall use the permutation test to estimate the null distribution of F despite its higher computational demand.

		$\alpha = 0.05$				$\alpha = 0.01$			
ϵ	Calibration	df	CV	AICC	h_s	df	CV	AICC	h_s
0.00	χ^2	0.090	0.075	0.070	0.055	0.030	0.030	0.025	0.010
	P. test	0.060	0.035	0.040	0.040	0.030	0.005	0.025	0.010
0.05	χ^2	0.220	0.195	0.120	0.090	0.090	0.085	0.060	0.030
	P. test	0.225	0.125	0.245	0.090	0.325	0.050	0.105	0.045
0.10	χ^2	0.320	0.350	0.265	0.145	0.210	0.220	0.170	0.055
	P. test	0.450	0.140	0.240	0.235	0.125	0.085	0.165	0.035
0.15	χ^2	0.555	0.535	0.460	0.360	0.385	0.400	0.330	0.230
	P. test	0.780	0.385	0.515	0.440	0.570	0.155	0.365	0.250
0.20	χ^2	0.795	0.760	0.730	0.535	0.685	0.675	0.630	0.395
	P. test	0.705	0.590	0.680	0.460	0.585	0.325	0.620	0.300
0.25	χ^2	0.965	0.965	0.960	0.930	0.955	0.960	0.955	0.885
	P. test	0.990	0.950	0.980	0.960	0.970	0.870	0.955	0.895
0.30	χ^2	0.985	0.965	0.965	0.945	0.975	0.965	0.965	0.930
	P. test	0.980	0.945	0.965	0.950	0.960	0.880	0.965	0.920
0.35	χ^2	1.000	1.000	1.000	0.990	1.000	1.000	1.000	0.985
	P. test	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990
0.40	χ^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	P. test	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.45	χ^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	P. test	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
0.50	χ^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	P. test	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE 6.1: Probability of rejecting separability for STPP1 (6.4.1) at significance levels $\alpha = 0.05$ and $\alpha = 0.01$. Comparison of calibration methods, χ^2: χ^2 approximation of the null distribution, P. test: permutation test. The kernel log-ratio function (6.2.3) was computed using the LSCV optimal bandwidth. Bandwidth selectors for the kernel regression function (6.3.1) in columns.

		$\alpha = 0.05$				$\alpha = 0.01$			
ϵ	Test	df	CV	AICC	h_s	df	CV	AICC	h_s
0.00	χ^2	0.110	0.130	0.100	0.075	0.055	0.075	0.050	0.020
	P. test	0.055	0.065	0.070	0.035	0.030	0.020	0.025	0.005
0.05	χ^2	0.100	0.135	0.090	0.065	0.020	0.050	0.025	0.025
	P. test	0.190	0.075	0.115	0.075	0.050	0.005	0.055	0.040
0.10	χ^2	0.235	0.440	0.400	0.390	0.085	0.150	0.130	0.135
	P. test	0.365	0.125	0.355	0.415	0.200	0.040	0.150	0.170
0.15	χ^2	0.600	0.695	0.660	0.640	0.415	0.515	0.510	0.495
	P. test	0.690	0.380	0.690	0.725	0.510	0.055	0.575	0.625
0.20	χ^2	0.780	0.850	0.875	0.860	0.685	0.760	0.750	0.725
	P. test	0.855	0.635	0.905	0.875	0.795	0.150	0.820	0.785
0.25	χ^2	0.930	0.945	0.950	0.930	0.870	0.915	0.915	0.905
	P. test	0.785	0.640	0.930	0.920	0.660	0.180	0.810	0.825
0.30	χ^2	0.955	0.975	0.970	0.965	0.930	0.940	0.945	0.945
	P. test	0.970	0.920	0.965	0.970	0.955	0.625	0.950	0.955
0.35	χ^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
	P. test	1.000	1.000	1.000	1.000	1.000	0.995	1.000	0.995
0.40	χ^2	0.995	0.995	0.990	0.995	0.990	0.990	0.990	0.990
	P. test	0.995	0.995	0.995	1.000	0.995	0.985	0.990	0.990
0.45	χ^2	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	P. test	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000
0.50	χ^2	0.995	1.000	1.000	1.000	0.995	0.995	0.995	0.995
	P. test	0.995	0.995	1.000	1.000	0.995	0.995	0.995	0.995

TABLE 6.2: Probability of rejecting separability for STPP2 (6.4.2) at significance levels $\alpha = 0.05$ and $\alpha = 0.01$. Comparison of calibration methods, χ^2: χ^2 approximation of the null distribution, P. test: permutation test. The kernel log-ratio function (6.2.3) was obtained using the LSCV-optimal bandwidth. Bandwidth selector for the kernel regression function (6.3.1) in columns.

6.4.2 Effect of the bandwidth parameters on the test

The separability test depends on two smoothing parameters: (i) the diagonal bandwidth matrix used to estimate $\rho(x, t)$ and (ii) the bandwidth parameter used to obtain the kernel regression function $\hat{m}(x)$. We have used different bandwidth selectors in both stages to check their effects on the performance of the test.

On one hand, the log-ratio function was estimated using the bandwidth minimizing the LSCV measure defined above (6.2.11), whose performance for the log-ratio estimation has not been tested yet; the 3D plug-in diagonal bandwidth for $\hat{f}(x, t)$ (Duong and Hazelton, 2003) taking into account the accuracy of plug-in bandwidth matrices to estimate the first-order intensity of spatial point processes (see Chapter 3); and the maximum smoothing parameter (Terrel, 1990) for density functions, which was used in Davies

et al. (2011) to estimate the spatial relative risk function. Table 6.3 shows that the LSCV bandwidth selector provides larger bandwidths than the plug-in and maximum smoothing selectors for the spatio-temporal densities.

On the other hand, to test whether the bandwidth used in the kernel regression (6.3.1) affects the performance of the F-test, we conducted this step using the spatial bandwidth in $\hat{\rho}(x,t)$, h_s, as well as the three bandwidth selectors currently available in the *sm* package (Bowman and Azzalini, 2014): approximated degrees of freedom of errors (df), least-squares cross-validation (CV), and the AICC-based bandwidth (AICC).

For separable patterns with spatio-temporal intensity functions (6.4.1), Table 6.4 ($\epsilon = 0$), and Figure 6.1 show that the plug-in and maximum smoothing bandwidths selectors provided large type I errors in the separability test, i.e. both bandwidths undersmoothed the log-ratio function and led to identify random variability in the separable point process as departure from separability. The probabilities of rejecting separability using the LSCV bandwidth selector were close to the nominal significance level, α. In agreement with these results we analyzed the effect of the bandwidth used to compute $\hat{m}(x)$ on the F-test using LSCV bandwidths in the kernel log-ratio function. Table 6.4 and Figure 6.2 report a good performance of the separability test with the four bandwidth selectors, as the type I errors under $\mathcal{H}_0$ were close to the nominal significance level, and the power of the test increased gradually with the departure from separability. The performance of the *df*-bandwidth was slightly worse than the performance of the other three procedures, as this method provided type I errors slightly larger than α, and led to higher probabilities of rejecting separability for $\epsilon = 0.15$ than for $\epsilon = 0.20$,

Table 6.5 and Figures 6.3, 6.4 and C.10, show the performance of the separability test for simulated patterns with spatio-temporal intensity function (6.4.2). As above, the plug-in and maximum smoothing bandwidths led to high probabilities of type I error for separable point patterns, while the LSCV bandwidth provided empirical p-values close to the nominal significance level. When we use the LSCV bandwidth in the kernel log-ratio function, the test had a good performance with the four bandwidth parameter in $\hat{m}(x)$, the type I errors under $\mathcal{H}_0$ were close to the nominal significance level, and the power of the test increased with the departure from separability. Figure 6.4 shows that the CV-bandwidth provide a slightly better approximation of the nominal significance level when $\epsilon = 0$.

These results confirm the good performance of the F-test with LSCV bandwidth selector in the kernel log-ratio function, while the smoothing parameter used in the regression barely affects the performance of the test.

		STPP1			STPP2		
ϵ	Selector	h_s		h_t	h_s		h_t
0.00	LSCV	0.75	0.57	0.12	0.75	0.75	0.11
	Max smooth	0.09	0.09	0.15	0.12	0.12	0.15
	Plug.in	0.07	0.07	0.08	0.09	0.09	0.09
0.05	LSCV	0.43	0.42	0.07	0.75	0.68	0.08
	Max smooth	0.09	0.09	0.15	0.12	0.12	0.15
	Plug.in	0.07	0.07	0.08	0.09	0.09	0.09
0.10	LSCV	0.62	0.75	0.09	0.75	0.61	0.11
	Max smooth	0.09	0.09	0.14	0.11	0.11	0.15
	Plug.in	0.07	0.07	0.08	0.09	0.09	0.09
0.15	LSCV	0.75	0.75	0.27	0.36	0.43	0.06
	Max smooth	0.09	0.09	0.14	0.11	0.11	0.15
	Plug.in	0.06	0.06	0.08	0.08	0.08	0.09
0.20	LSCV	0.75	0.75	0.19	0.56	0.59	0.08
	Max smooth	0.09	0.09	0.14	0.11	0.11	0.15
	Plug.in	0.06	0.06	0.07	0.08	0.08	0.09
0.25	LSCV	0.53	0.61	0.07	0.58	0.64	0.08
	Max smooth	0.08	0.08	0.14	0.11	0.11	0.15
	Plug.in	0.06	0.06	0.07	0.08	0.08	0.09
0.30	LSCV	0.41	0.39	0.28	0.54	0.66	0.10
	Max smooth	0.08	0.08	0.14	0.10	0.10	0.15
	Plug.in	0.05	0.05	0.06	0.08	0.08	0.09
0.35	LSCV	0.52	0.56	0.22	0.50	0.48	0.45
	Max smooth	0.08	0.08	0.13	0.10	0.10	0.15
	Plug.in	0.05	0.05	0.06	0.08	0.08	0.09
0.40	LSCV	0.75	0.67	0.07	0.54	0.50	0.33
	Max smooth	0.07	0.07	0.13	0.10	0.10	0.15
	Plug.in	0.05	0.05	0.06	0.08	0.08	0.09
0.45	LSCV	0.75	0.31	0.35	0.22	0.23	0.34
	Max smooth	0.07	0.07	0.12	0.10	0.10	0.15
	Plug.in	0.05	0.05	0.05	0.08	0.08	0.09
0.50	LSCV	0.30	0.33	0.31	0.29	0.26	0.22
	Max smooth	0.06	0.06	0.11	0.10	0.10	0.15
	Plug.in	0.04	0.04	0.05	0.07	0.07	0.08

TABLE 6.3: Comparison of bandwidth selectors for the spatio-temporal point processes with intensity functions (6.4.1) and (6.4.2)

		$\alpha = 0.05$				$\alpha = 0.01$			
ϵ	Selector	df	CV	AICC	h_s	df	CV	AICC	h_s
0.00	LSCV	0.060	0.035	0.040	0.040	0.030	0.005	0.025	0.010
	Max smooth	0.465	0.190	0.490	0.565	0.320	0.100	0.425	0.465
	Plug-in	0.755	0.440	0.725	0.840	0.645	0.295	0.660	0.790
0.05	LSCV	0.225	0.125	0.245	0.090	0.325	0.050	0.105	0.045
	Max smooth	0.630	0.355	0.585	0.750	0.555	0.230	0.520	0.645
	Plug-in	0.900	0.595	0.840	0.975	0.845	0.415	0.805	0.885
0.10	LSCV	0.450	0.140	0.240	0.235	0.125	0.085	0.165	0.035
	Max smooth	0.670	0.485	0.690	0.685	0.550	0.245	0.570	0.560
	Plug-in	0.885	0.565	0.880	0.875	0.825	0.400	0.835	0.790
0.15	LSCV	0.780	0.385	0.515	0.440	0.570	0.155	0.365	0.250
	Max smooth	0.940	0.760	0.885	0.905	0.830	0.510	0.735	0.780
	Plug-in	0.995	0.845	0.970	0.990	0.990	0.635	0.945	0.980
0.20	LSCV	0.705	0.590	0.680	0.460	0.585	0.325	0.620	0.300
	Max smooth	0.950	0.790	0.930	0.935	0.870	0.605	0.895	0.880
	Plug-in	0.945	0.845	0.965	0.795	0.845	0.645	0.485	0.535
0.25	LSCV	0.990	0.950	0.980	0.960	0.970	0.870	0.955	0.895
	Max smooth	0.975	0.955	0.985	0.865	0.925	0.835	0.955	0.780
	Plug-in	0.950	0.970	0.900	0.710	0.750	0.115	0.270	0.450
0.30	LSCV	0.980	0.945	0.965	0.950	0.960	0.880	0.965	0.920
	Max smooth	0.985	0.960	1.000	0.995	0.980	0.840	0.985	0.985
	Plug-in	0.990	0.970	0.985	0.925	0.900	0.735	0.815	0.810
0.35	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.990
	Max smooth	0.995	0.980	1.000	0.935	0.970	0.690	0.885	0.805
	Plug-in	0.995	1.000	0.995	0.990	0.995	0.900	0.975	0.955
0.40	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Max smooth	1.000	0.995	1.000	1.000	1.000	0.990	1.000	1.000
	Plug-in	0.980	0.995	0.885	0.945	0.955	0.415	0.510	0.685
0.45	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Max smooth	1.000	0.995	1.000	0.995	1.000	0.975	1.000	0.995
	Plug-in	1.000	1.000	1.000	0.995	0.990	0.990	0.915	0.980
0.50	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Max smooth	1.000	1.000	1.000	0.990	0.985	0.860	0.905	0.925
	Plug-in	1.000	1.000	1.000	0.985	1.000	0.990	1.000	0.960

TABLE 6.4: Probability of rejecting separability for the spatio-temporal point processes with first-order intensity (6.4.1). Null distribution of F estimated with $B = 500$ realizations of the permutation test. Bandwidth selector for the kernel log-ratio function (6.2.3) in rows, and bandwidth selector for the kernel regression function (6.3.1) in columns.

FIGURE 6.1: Effect of the bandwidth parameter used in $\hat{\rho}(x,t)$ on the probability of accepting separability for STPP1. Each plot corresponds to a bandwidth parameter in the regression step, df: approximated degrees of freedom of the errors; CV: least squares cross-validation, AICC: AICC-based bandwidth. h_s; spatial bandwidth in the log-ratio estimator

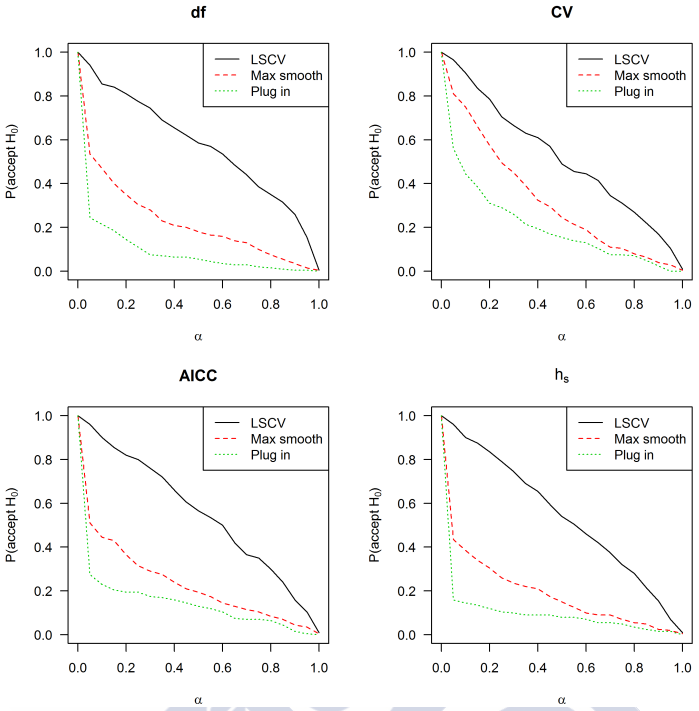
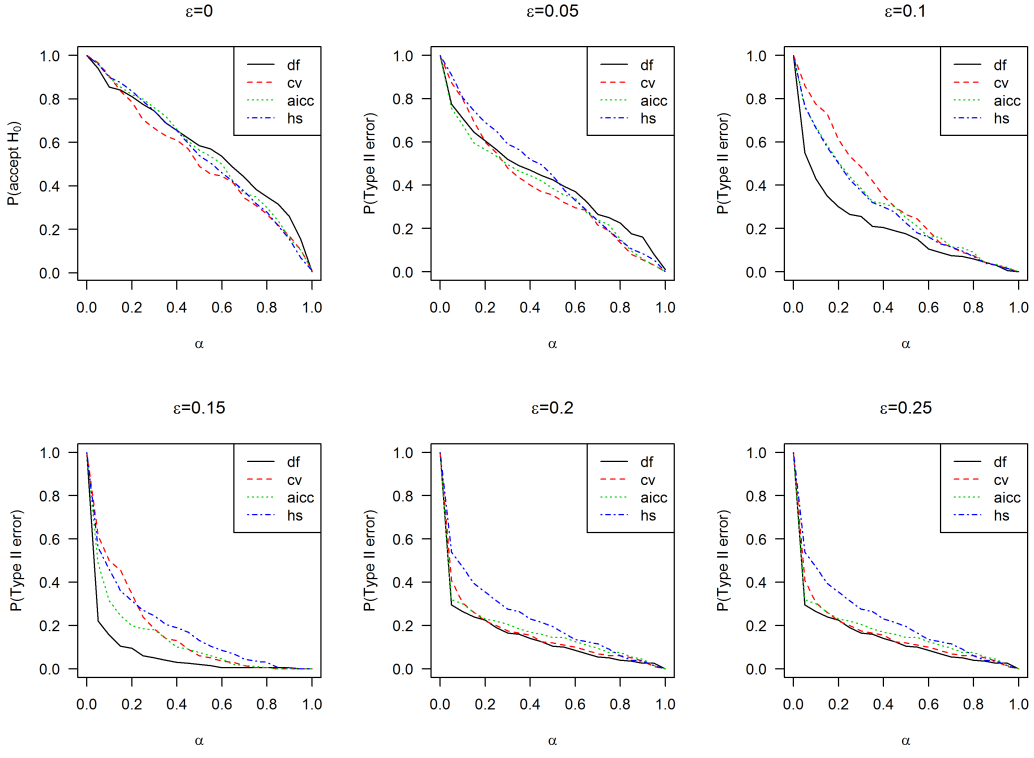


FIGURE 6.2: Effect of the bandwidth parameter used in $\hat{m}(x)$ on the separability test for STPP1. Probability of acceptance for separable patterns, and type II error for nonseparable patterns. LSCV bandwidth in the kernel log-ratio function.



		$\alpha = 0.05$				$\alpha = 0.01$			
ϵ	Selector	df	CV	AICC	h_s	df	CV	AICC	h_s
	LSCV	0.055	0.065	0.070	0.035	0.030	0.020	0.025	0.005
0.00	Max smooth	0.570	0.315	0.425	0.605	0.405	0.070	0.320	0.440
	Plug-in	0.755	0.400	0.625	0.815	0.555	0.185	0.450	0.620
	LSCV	0.190	0.075	0.115	0.075	0.050	0.005	0.055	0.040
0.05	Max smooth	0.480	0.265	0.360	0.680	0.370	0.130	0.270	0.505
	Plug-in	0.420	0.200	0.425	0.410	0.260	0.110	0.345	0.245
	LSCV	0.365	0.125	0.355	0.415	0.200	0.040	0.150	0.170
0.10	Max smooth	0.385	0.310	0.450	0.280	0.215	0.095	0.320	0.120
	Plug-in	0.445	0.490	0.390	0.380	0.175	0.070	0.040	0.095
	LSCV	0.690	0.380	0.690	0.725	0.510	0.055	0.575	0.625
0.15	Max smooth	0.515	0.690	0.590	0.295	0.305	0.320	0.025	0.090
	Plug-in	0.575	0.415	0.680	0.415	0.450	0.210	0.505	0.325
	LSCV	0.855	0.635	0.905	0.875	0.795	0.150	0.820	0.785
0.20	Max smooth	0.850	0.755	0.950	0.775	0.620	0.375	0.620	0.470
	Plug-in	0.890	0.655	0.915	0.820	0.815	0.460	0.765	0.735
	LSCV	0.785	0.640	0.930	0.920	0.660	0.180	0.810	0.825
0.25	Max smooth	0.955	0.935	0.930	0.815	0.695	0.510	0.385	0.500
	Plug-in	0.595	0.805	0.080	0.330	0.305	0.265	0.020	0.055
	LSCV	0.970	0.920	0.965	0.970	0.955	0.625	0.950	0.955
0.30	Max smooth	0.990	0.985	0.990	0.890	0.795	0.920	0.480	0.690
	Plug-in	0.590	0.810	0.410	0.400	0.250	0.470	0.035	0.045
	LSCV	1.000	1.000	1.000	1.000	1.000	0.995	1.000	0.995
0.35	Max smooth	0.925	0.870	1.000	0.845	0.845	0.630	0.910	0.770
	Plug-in	0.725	0.710	0.760	0.495	0.460	0.340	0.405	0.405
	LSCV	0.995	0.995	0.995	1.000	0.995	0.985	0.990	0.990
0.40	Max smooth	0.905	0.890	0.995	0.805	0.845	0.605	0.920	0.675
	Plug-in	0.405	0.880	0.005	0.360	0.290	0.570	0.005	0.005
	LSCV	1.000	1.000	1.000	1.000	1.000	0.995	1.000	1.000
0.45	Max smooth	0.930	0.825	1.000	0.715	0.840	0.645	0.955	0.610
	Plug-in	0.600	0.610	0.435	0.440	0.430	0.410	0.390	0.405
	LSCV	0.995	0.995	1.000	1.000	0.995	0.995	0.995	0.995
0.50	Max smooth	0.990	0.995	0.995	0.870	0.930	0.555	0.550	0.600
	Plug-in	0.870	0.780	0.990	0.675	0.790	0.550	0.930	0.585

TABLE 6.5: Probability of rejecting separability for STPP2. Null distribution of F estimated with $B = 500$ realizations of the permutation test. Bandwidth selector for the kernel log-ratio function (6.2.3) in rows, and bandwidth selector for the kernel regression function (6.3.1) in columns.

FIGURE 6.3: Effect of the bandwidth parameter used in $\hat{\rho}(x,t)$ on the probability of accepting separability for STPP2. Each plot corresponds to a bandwidth parameter in the regression step, df: approximated degrees of freedom of the errors; CV: least squares cross-validation, AICC: AICC-based bandwidth. h_s; spatial bandwidth in the log-ratio estimator

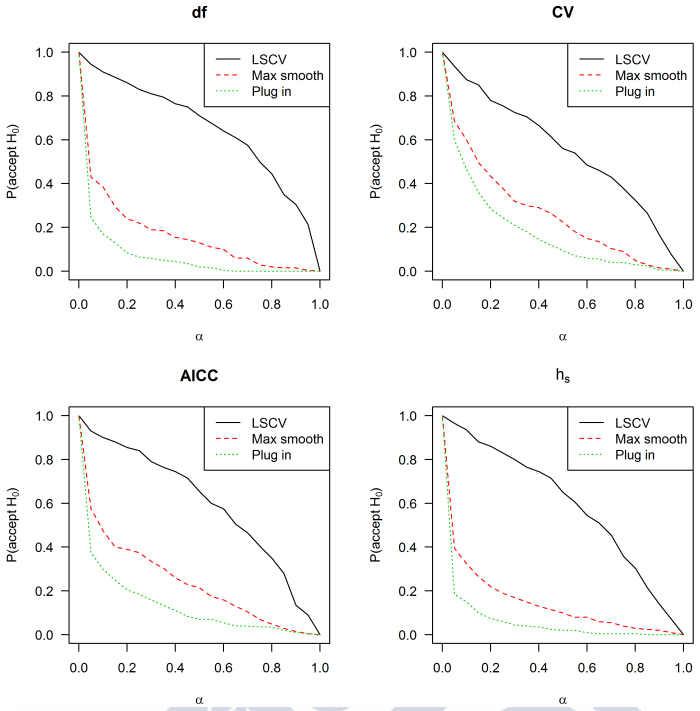
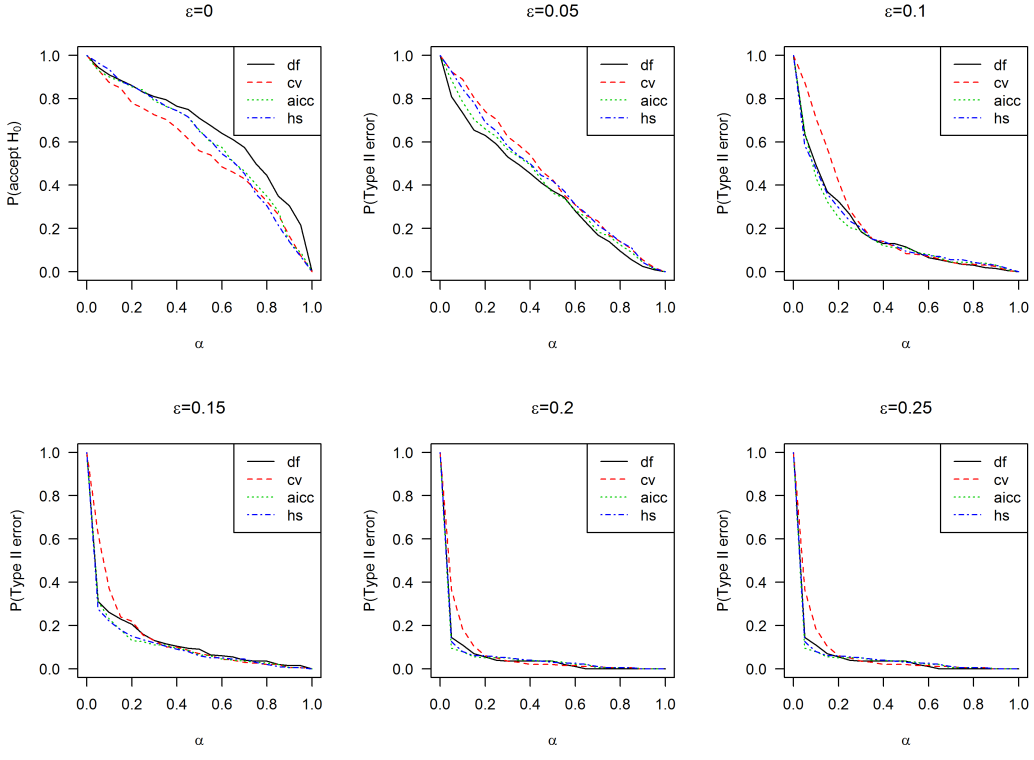


FIGURE 6.4: Effect of the bandwidth parameter used in $\hat{m}(x)$ on the separability test for STPP2.Probability of acceptance for separable patterns, and type II error for nonseparable patterns. LSCV bandwidth in the kernel log-ratio function.



6.4.3 Comparison with previous separability tests

In this section we compare the performance of the nonparametric separability test introduced in this work with those proposed by Schoenberg (2004) and Díaz-Avalos et al. (2013) (see details in Section 2.4.3). Taking into account the results obtained above, in this simulation study the test statistic, $\hat{F}$, was obtained using LSCV bandwidths in $\hat{\rho}(x,t)$ and CV bandwidth in $\hat{m}(x)$, and random permutations were used to estimate the null distribution of F. The nonparametric test based on F was compared with tests based on the Cramer-von-Mises type statistic, S_3 (2.4.15), and the log-likelihood test, S_4 (2.4.16), introduced by Schoenberg (2004). These statistics were obtained using Gaussian kernels with bandwidth 1/20 times the range of each dimension ($h_s = h_t = 0.05$), to estimate the separable, $\lambda^S(x,t)$, and nonseparable, $\lambda^{NS}(x,t)$, spatio-temporal intensity functions. We also compared the F-test with the Kullback-Leibler, KL (2.4.17), and Hellinger, H (2.4.18), discrepancy measures introduced by Díaz-Avalos et al. (2013). These measures were obtained using Gaussian kernels with bandwidth $h = 0.68n^{-3/10}$ ($h_s = h_t \approx 0.105$, for $m = 500$), which allows a reasonable degree of smoothing for different number of dimensions, to estimate the separable and nonseparable intensity functions.

To conduct the tests based on S_3, S_4, KL, and H we obtained the separable, $\hat{\lambda}^S(x,t)$, and nonseparable, $\hat{\lambda}^{NS}(x,t)$, kernel estimators of the spatio-temporal intensity functions for the observed pattern and computed the corresponding test statistic, $\hat{u}$. We also computed the test statistic $\{u_b^*, b = 1,\ldots,B\}$ for $B = 500$ realizations of separable point processes with intensity function $\hat{\lambda}^S(x,t)$. The empirical p-value of the test is the proportion of simulated u^ss larger than $\hat{u}$. As in the previos simulation studies,the performance of the test was compared on 200 simulations of STPP1 (6.4.1) and STPP2 (6.4.2). For any significance level, α, the proportion of times that the test statistic, u_1, for the observed pattern is higher than the corresponding $100(1-\alpha)$-est quantile under H_0 indicates the probability of rejecting the null hypothesis.

For STPP1, Table 6.6 reports probabilities of rejecting $\mathcal{H}_0$ lower than the nominal significance levels when $\epsilon = 0$ with the four benchmark tests. Although as α increases the log-likelihood test provided high probabilities of type I error, while the probabilities of rejecting the null hypothesis with the test based on S_3, KL and H were below the nominal significance level (Figure 6.5). As observed in the previous simulation study, the F-test provided a good approximation of the nominal significance level. For nonseparable point processes, $\epsilon > 0$, we observe a fast increase in the power of S_3, KL and H as the observed patterns deviates from the null hypothesis. The power of the F-test

		Permutation test				Previous tests			
ϵ	α	df	CV	AICC	h_s	S_3	S_4	KL	H
0.00	0.01	0.030	0.005	0.025	0.010	0.005	<0.005	0.005	0.005
	0.05	0.060	0.035	0.040	0.040	0.020	0.030	0.005	0.005
0.05	0.01	0.325	0.050	0.105	0.045	0.180	0.005	0.035	0.035
	0.05	0.450	0.140	0.240	0.235	0.410	0.075	0.120	0.120
0.10	0.01	0.125	0.085	0.165	0.035	0.745	0.070	0.475	0.560
	0.05	0.225	0.125	0.245	0.090	0.900	0.275	0.835	0.840
0.15	0.01	0.570	0.155	0.365	0.250	0.985	0.185	0.810	0.835
	0.05	0.780	0.385	0.515	0.440	0.995	0.505	0.960	0.975
0.20	0.01	0.585	0.325	0.620	0.300	1.000	0.640	0.975	0.990
	0.05	0.705	0.590	0.680	0.460	1.000	0.715	1.000	1.000
0.25	0.01	0.970	0.870	0.955	0.895	1.000	0.525	0.975	0.995
	0.05	0.990	0.950	0.980	0.960	1.000	0.760	1.000	1.000
0.30	0.01	0.960	0.880	0.965	0.920	1.000	0.780	0.995	0.995
	0.05	0.980	0.945	0.965	0.950	1.000	0.860	1.000	1.000
0.35	0.01	1.000	1.000	1.000	0.990	1.000	0.920	0.995	1.000
	0.05	1.000	1.000	1.000	1.000	1.000	0.940	1.000	1.000
0.40	0.01	1.000	1.000	1.000	1.000	1.000	0.935	0.980	0.990
	0.05	1.000	1.000	1.000	1.000	1.000	0.950	1.000	1.000
0.45	0.01	1.000	1.000	1.000	1.000	1.000	0.965	0.990	0.990
	0.05	1.000	1.000	1.000	1.000	1.000	0.970	0.990	1.000
0.50	0.01	1.000	1.000	1.000	1.000	1.000	0.980	0.810	0.910
	0.05	1.000	1.000	1.000	1.000	1.000	0.995	0.980	0.980

TABLE 6.6: Comparison of the F-test with previous tests for STPP1. Probability of rejecting the null hypothesis at significance levels $\alpha = 0.05$ and $\alpha = 0.01$.

and S_4 also increased with the departure from separability but at a slower rate. Thus, in this case, S_4 had a poor performance, while the other benchmark tests where more powerful than the F-test (see also Figure C.12).

For STPP2, S_3 provided type I errors close to the nominal significance level, while the probabilities of rejecting separability with the remainder benchmark tests are lower than α (Table 6.6). The F-test provided good approximations of the nominal significance level. Figure 6.6 show that the F-test is more powerful than the benchmark criteria for $\epsilon < 0.25$, when the tests based on KL and H outperformed our tests. Our results also highlight the poor performance of S_4 for these point processes (see also Figure C.12).

FIGURE 6.5: Comparison of the F-test with previous separability tests for STPP1. Probability of accepting $\mathcal{H}_0$ for separable point patterns, and type II error for nonseparable point patterns. F: F-test with LSCV bandwidth in $\hat{\rho}(x,t)$ and AICC bandwidth in $\hat{m}(x)$. S_3 and S_4 by Schoenberg (2004). KL and H discrepancy measures by Díaz-Avalos et al. (2013).

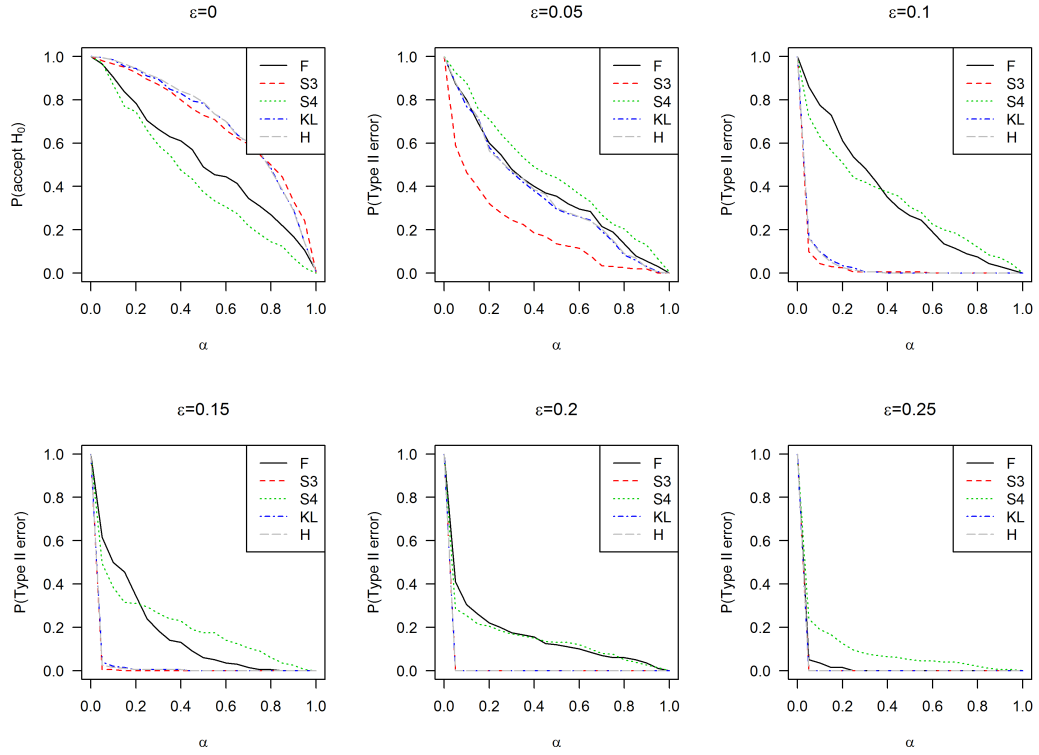
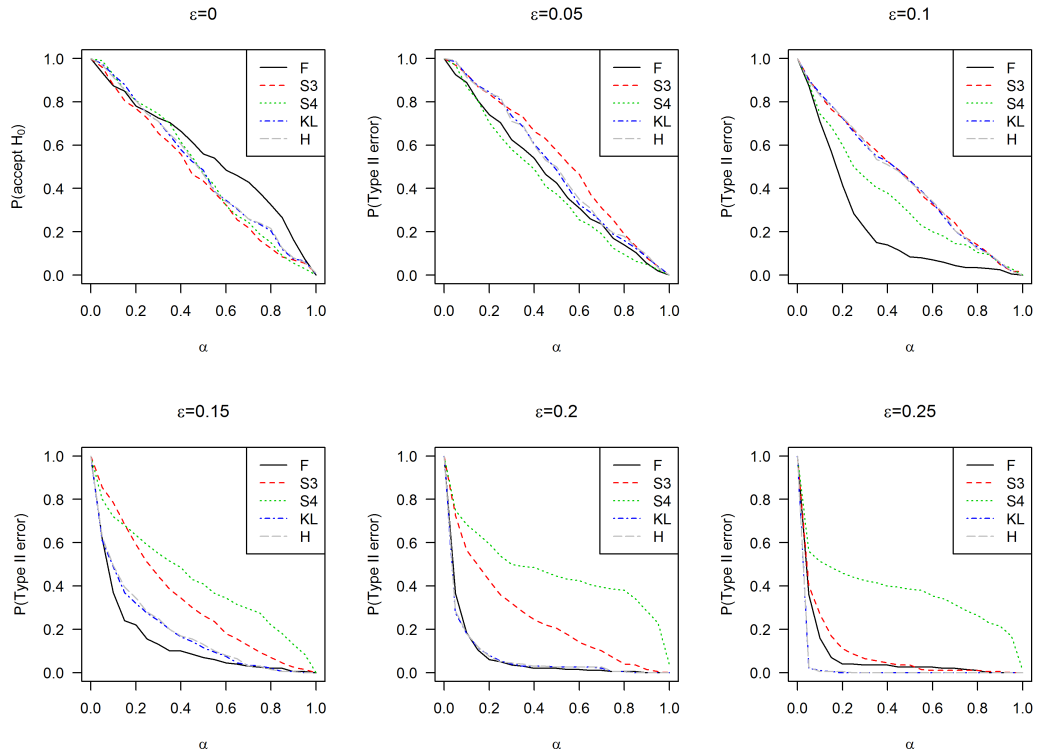


FIGURE 6.6: Comparison of The F-test with previous separability tests for STPP2. See details in the caption of Figure 6.5.



		Permutation test				Previous tests			
ϵ	α	df	CV	AICC	h_s	S_3	S_4	KL	H
0.00	0.01	0.030	0.020	0.025	0.005	0.015	0.005	<0.005	<0.005
	0.05	0.055	0.065	0.070	0.035	0.040	0.010	0.025	0.030
0.05	0.01	0.050	0.005	0.055	0.040	0.005	<0.005	<0.005	<0.005
	0.05	0.190	0.075	0.115	0.075	0.030	0.030	0.015	0.010
0.10	0.01	0.200	0.040	0.150	0.170	0.035	0.030	0.035	0.050
	0.05	0.365	0.125	0.355	0.415	0.095	0.115	0.100	0.105
0.15	0.01	0.510	0.055	0.575	0.625	0.065	0.085	0.190	0.170
	0.05	0.690	0.380	0.690	0.725	0.145	0.200	0.390	0.375
0.20	0.01	0.795	0.150	0.820	0.785	0.060	0.060	0.425	0.435
	0.05	0.855	0.635	0.905	0.875	0.275	0.250	0.725	0.715
0.25	0.01	0.660	0.180	0.810	0.825	0.385	0.335	0.890	0.890
	0.05	0.785	0.640	0.930	0.920	0.600	0.440	0.980	0.975
0.30	0.01	0.955	0.625	0.950	0.955	0.505	0.290	0.965	0.955
	0.05	0.970	0.920	0.965	0.970	0.745	0.425	0.995	0.995
0.35	0.01	1.000	0.995	1.000	0.995	0.890	0.472	0.995	0.995
	0.05	1.000	1.000	1.000	1.000	0.975	0.518	0.995	0.995
0.40	0.01	0.995	0.985	0.990	0.990	1.000	0.520	1.000	1.000
	0.05	0.995	0.995	0.995	1.000	1.000	0.561	1.000	1.000
0.45	0.01	1.000	0.995	1.000	1.000	1.000	0.545	1.000	0.995
	0.05	1.000	1.000	1.000	1.000	1.000	0.592	1.000	1.000
0.50	0.01	0.995	0.995	0.995	0.995	1.000	0.553	1.000	1.000
	0.05	0.995	0.995	1.000	1.000	1.000	0.580	1.000	1.000

TABLE 6.7: Comparison of the F-test with previous separability tests for STPP2. Probability of rejecting the null hypothesis at significance levels $\alpha = 0.05$ and $\alpha = 0.01$.

6.5 Testing separability in the spatio-temporal pattern of wildfires registered in Galicia

The wildfires dataset contains the spatial locations and time of occurrence for the ignition point of wildfires registered in Galicia. We have applied the nonparametric separability test proposed above to check whether the spatial distribution of wildfires registered during 2006 remain constant over time. We tested the separability assumption on the unmarked pattern (see the marginal spatial and temporal point processes in Figure 6.7) and on the spatio-temporal point patterns classified by burned area (Figure 6.8) and cause (Figure 6.9). Taking into account the results of the simulation study in Section 6.4.2, the bandwidth matrices in $\hat{\rho}(x,t)$ were selected by LSCV (6.2.11), and we used the four bandwidth selectors to obtain $\hat{m}(x)$. The test was calibrated by $B = 500$ realizations of the permutation test.

In Figure 6.7 we can see that only a few wildfires were registered from October onwards.

Thus, in order to avoid the negative effects of data sparseness on the kernel log-ratio function and, consequently on the power of the separability test, we have tested the separability of all the wildfire patterns in the period January-September. Likewise, most of the large fires (6.8), natural fires, and reproductions (6.9) were observed during summer. Thus, for these patterns we have applied the separability test in the whole period and shortening the temporal interval according to the period where fires were observed, to check the effect of temporal sparseness on the performance of the test.

Table 6.8 shows the bandwidth parameters used to estimate the log-ratio function, and the p-values of the separability test obtained with the four bandwidths in the regression step. We should first note that, as expected, kernel smoothing of the log-ratio function requires larger bandwidths than kernel intensity estimation (see Tables 5.2 and 5.3). The test statistic rejects the null hypothesis for all the patterns analyzed, i.e. their spatial distribution varies throughout the year. The results obtained for the unmarked spatio-temporal pattern agree with that provided in Section 4.4 by the nonparametric comparison between months.

Table 6.8 shows that data sparseness did not affect the results of the test for large fires, natural fires and negligence, whose events where concentrated on a shorter temporal interval. However data sparseness may affect the LSCV bandwidth selector, which provided larger spatial bandwidths for the whole period.

FIGURE 6.7: Spatial and temporal patterns of wildfires registered in Galicia during 2006

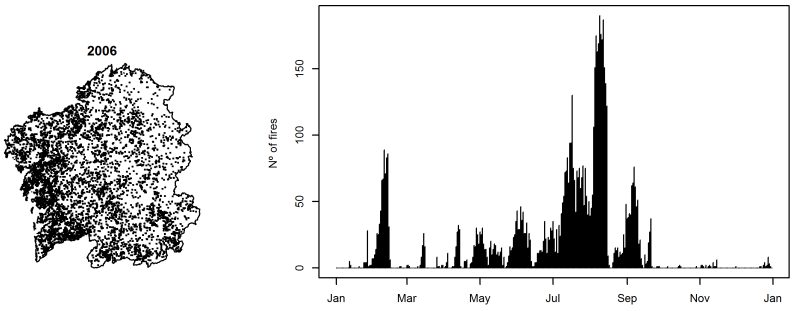
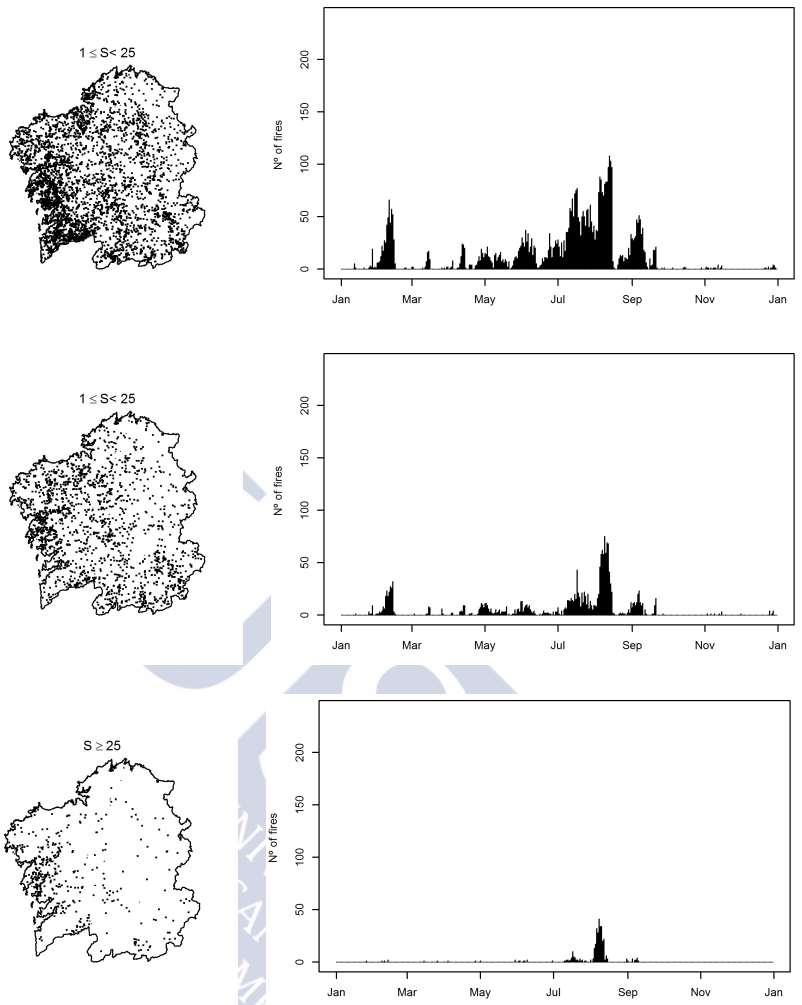


FIGURE 6.8: Spatial and temporal patterns of wildfires registered in Galicia during 2006 by burned area

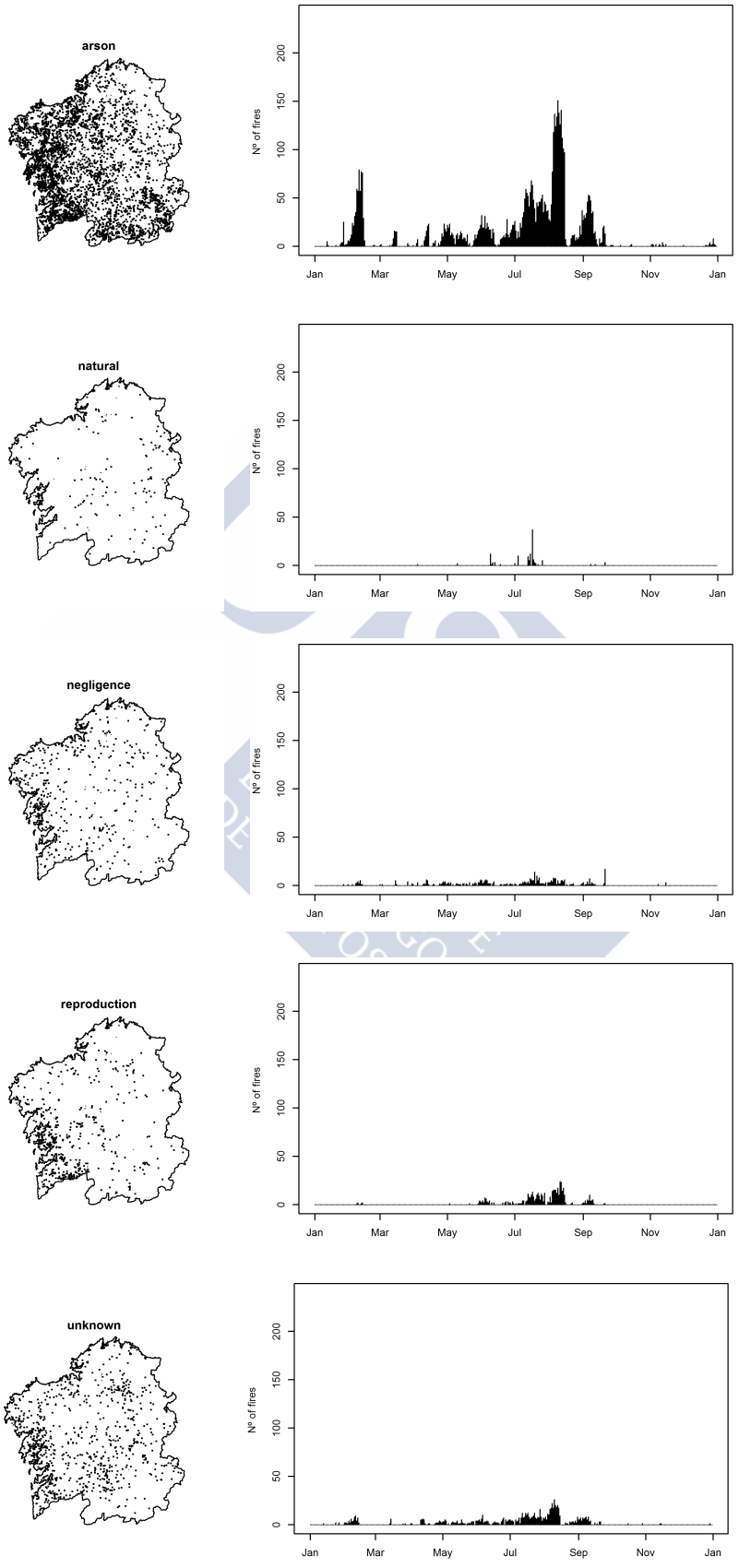


6.6 Conclusions

Testing the separability assumption should be among the first steps in the analysis of spatio-temporal point processes, due to its implication for the modeling of the spatio-temporal intensity. Up to date, this hypothesis has been tested by nonparametric tests calibrated through simulations of separable point processes.

In a separable spatio-temporal point process the risk of observing an event at time t is spatially invariant, i.e. the ratio between the spatio-temporal and spatial intensity functions does not depend on the spatial locations of events. In this chapter we develop a nonparametric separability test based on this property, which tests the dependence of the log-ratio function on the spatial location through the nonparametric no-effect test proposed by Bowman and Azzalini (1997). In order to implement the test, we developed

FIGURE 6.9: Spatial and temporal patterns of wildfires registered in Galicia during 2006 by cause



			Bandwidth $\hat{\rho}(x,t)$			p-value F-test			
		n	h_s		h_t	df	CV	AICC	h_s
Wildfires		6851	69.93	105.62	6.78	< 0.002	< 0.002	< 0.002	< 0.002
Size	$S < 1$	4764	50.39	102.92	9.68	< 0.002	< 0.002	< 0.002	< 0.002
	$1 \leq S < 25$	1750	25.60	66.56	11.72	< 0.002	< 0.002	< 0.002	< 0.002
	$S \geq 25$	337	76.55	124.04	6.78	0.00	< 0.002	< 0.002	< 0.002
	$S \geq 25$ (Jn-S)	325	27.42	40.80	39.07	0.00	< 0.002	< 0.002	< 0.002
Cause	arson	5046	39.84	107.40	10.11	< 0.002	< 0.002	< 0.002	< 0.002
	natural	120	100.82	120.91	6.13	< 0.002	< 0.002	< 0.002	< 0.002
	natural (Jn,.Jl)	112	74.55	105.84	3.52	< 0.002	< 0.002	< 0.002	< 0.002
	negligence	396	67.41	55.16	12.89	< 0.002	< 0.002	< 0.002	< 0.002
	reproduction	475	42.29	107.39	6.70	< 0.002	< 0.002	< 0.002	< 0.002
	rep (My-S)	459	37.85	46.60	5.03	< 0.002	< 0.002	< 0.002	< 0.002
	unknown	814	31.59	39.47	11.54	< 0.002	< 0.002	< 0.002	< 0.002

TABLE 6.8: Nonparametric separability test for the wildfires registered during 2006 (January-February) by size and cause. LSCV bandwidth for the kernel log-ratio function, and p-values obtained with the four alternative bandwidth selectors in the regression stage.

a kernel estimator of the log-ratio function, using the similarity between spatio-temporal relative risk function and the ratio between the spatio-temporal and the spatial intensity functions, and a least-squares cross-validation bandwidth selection procedure for this kernel estimator.

The results of the simulation study conducted to analyze the performance of the test show that a permutation test provides a better calibration of the test than a χ^2 approximation of the null distribution. The analysis of the effect of the bandwidth vector on the performance of the test, reported a good approximations of the nominal significance level under the null hypothesis with the LSCV bandwidth, while the plug-in and maximum smmothing bandwidths undersmoothed the log-ratio function with the subsequent increase in the type I error. The separability test with LSCV bandwidth also performed well under the alternative hypothesis, as it was able to detect different ways of departure from separability and its power increased with the departure from separability.

Comparison with prior separability tests (Schoenberg, 2004; Díaz-Avalos et al., 2013), showed a better behavior of the F-test under the null hypothesis, and that is competitive with previous nonparametric tests under the alternative hypothesis. In particular, the test based on F was more powerful than the test based on S_4 for the two point processes simulated. These results support the validity of the nonparametric separability test introduced in this chapter.

In the application to real data the separability test detected departure from separability in all the spatio-temporal patterns analyzed. We have also seen how data sparseness affects the LSCV bandwidth selector. Thus, although in our case the two bandwidths led to the same results in the F-tests, given the evidence against separability, we recommend adjusting the temporal interval to the data in order to avoid the negative effects of data sparseness.

Chapter 7

Future research

7.1 Introduction

This chapter introduces some ideas for future contributions on different areas related with the scope of the thesis. Section 7.2 deals with the first-order characteristics of spatial point processes: Section 7.2.1 discusses the use of variable bandwidths in the consistent kernel estimator introduced in Chapter 3; in Section 7.2.2 we discuss some extensions of the nonparametric comparison of first-order intensities introduced in Chapter 4; and Section 7.2.3 addresses the development of goodness-of-fit tests. Section 7.3 discusses some ideas that should be taken into account to obtain accurate estimators of the spatio-temporal intensity function.

7.2 Spatial point processes

7.2.1 Adaptive kernel intensity estimator

The two main issues in kernel intensity estimation for spatial point processes are edge-effects and bandwidth selection. The edge-corrector $p_H(x) = \int_W k_H((x-y))dy$ guarantees that the kernel intensity estimator is asymptotically unbiased, but the bias of $\hat{\lambda}_{0,H}(x)$ near the boundary of W has order $O\left(tr(H)^{1/2}\right)$, while the bias in the interior of the observation domain is $O\left(tr(H)\right)$. Hazelton and Marshall (2009) introduced a linear boundary kernel with scalar bandwidth for bivariate densities with $O(h^2)$ bias near the boundary as in the interior of the observation domain. We can analyze the performance of the linear boundary kernels with bandwidth matrix for first-order intensity estimation.

In Chapter 3 we have seen that the kernel intensity estimator with plug-in bandwidth matrix has a reasonable performance even for highly inhomogeneous point processes. However, kernel intensity estimators with fixed bandwidth may not be well suited when dealing with highly inhomogeneous datasets. Barr and Schoenberg (2010) pointed out that kernel intensity estimators with fixed scalar bandwidth can be substantially biased when the intensity function is highly volatile and may have relatively large variance, specially in areas with low intensity. These authors introduced a Voronoi intensity estimator that outperforms kernel smoothing with fixed scalar bandwidth. The Voronoi intensity estimator is less biased, particularly in locations of low intensity surrounded by locations of high intensity, and may reduce the variability of the kernel estimator in areas with low intensity. Marshall and Hazelton (2010) introduced an adaptive kernel density estimator for bivariate distributions observed in a bounded domains, $W \in \mathbb{R}^2$, which can be also used to estimate the first-order intensity of inhomogeneous spatial point processes.

Let $\{x_1, \ldots, x_n\}$ be a random sample of the bivariate random variable with density $f(x)$. The adaptive kernel density estimator of f is

$$\hat{f}_{adv}(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{h_i} k\left(\frac{x - x_i}{h_i}\right) \tag{7.2.1}$$

where the kernel function, $k(\cdot)$, is a radially symmetric bivariate probability density function, and $h_i = h(x_i) = h_0 f(x_i)^{-1/2}$ is the variable bandwidth, with global bandwidth h_0. The adaptive bandwidth is large in areas with low density and small in areas with high density. The bias of $\hat{f}_{adv}(x)$ is $O\left(h_0^4\right)$ for any point in the interior of W, and the variance of $\hat{f}_{adv}(x)$ has the same order, $O\left(n^{-1}h_0^{-1}\right)$, as the variance of the kernel estimator with fixed bandwidth. As for bivariate kernel densities with fixed bandwidth, linear boundary kernels can be used to obtain a bias of order $O(h_0^2)$ in the boundary of the observation region.

Implementation of the adaptive estimator, 7.2.1, requires a pilot estimator of f to compute the variable bandwidth, $h_i = h_0 \hat{f}(x_i)^{-1/2}$. Marshall and Hazelton (2010) showed that using a fixed bandwidth estimator $\hat{f}_h(\cdot)$ with $h = O(n^{-1/6})$ and defining the local bandwidth for the adaptive estimator as $h_i = h_0 \hat{f}_h(x_i)$, where $h/h_0 \to 0$ as $n \to \infty$, the bias of the adaptive kernel density estimator with pilot density, $\hat{f}_h$, has the same order as the bias obtained using the target density, f.

The relationship between the density of event locations and the bivariate density function suggest using adaptive kernel estimators in the point process framework. However a correct implementation of this method require accurate selectors of the pilot, h, and global, h_0, bandwidths. Plug-in algorithms can be used to obtain the pilot bandwidth, but further research need to be conducted in order to estimate h_0. In Section 3.3.2 we have seen that the bootstrap resampling should be conducted with a bandwidth of order $O(h^{-1/8})$, i.e. asymptotically larger than the optimal bandwidth for the kernel estimator of $\lambda_0(x)$, as required for the global bandwidth in the adaptive kernel estimator. Therefore, we can check the effect of global bandwidths of order $O(h^{-1/8})$ on the performance of the adaptive kernel estimator of the density of event locations.

7.2.2 Extensions of the nonparametric comparison of first-order intensities

Chapter 4 provides a nonparametric test for pairwise comparison of first-order intensities. However we can be interested on testing the equality of the p first-order intensities in a multitype spatial point process. In this case our aim is to decide between the following hypothesis

$$\mathcal{H}_0 : \lambda_{01}(x), \ldots, \lambda_{0p}(x)$$
$$\mathcal{H}_1 : \lambda_{0i}(x) \neq \lambda_{0j}(x), \text{ for any } i \neq j$$

To this purpose we can extend the discrepancy measure T 4.2.2 and define

$$Q = \sum_{i=1}^{p-1} \sum_{j \geq i} \int_W \left(\lambda_{0i}(x) - \lambda_{0j}(x)\right)^2 dx \tag{7.2.2}$$

that yields the following test statistic

$$\hat{Q} = \sum_{i=1}^{p-1} \sum_{j \geq i} = \left(\hat{\psi}_{0,i} + \hat{\psi}_{0,j} - \left(\hat{\psi}_{0,ij} + \hat{\psi}_{0,ji}\right)\right) \tag{7.2.3}$$

We can also extend the bootstrap procedure introduced in Chapter 4 to the multitype framework to calibrate the null distribution of $\hat{Q}$. The design of this bootstrap calibration should be conducted in the future.

The T-test introduced in Chapter 4 allow us to analyze the global discrepancy between the first-order structure of two spatial point processes. We may also be interested on

measuring local discrepancies between first-order intensity functions, for instance we can wonder whether the risk of arson wildfires at a given location is higher than the risk of natural wildfires. Considering once again the equivalence between the bivariate density function and the density of event locations, we can extend the nonparametric test introduced by Duong (2013b) to measure local discrepancies between multivariate density functions:

$$\hat{U}(x) = \left(\hat{\lambda}_{01,H_1}(x) - \hat{\lambda}_{02,H_2}(x)\right)^2 \tag{7.2.4}$$

In the multivariate data framework, $U(x)$ is asymptotically chi-squared distributed under the null hypothesis. Further research should be conducted to establish the asymptotic null distribution of $U(x)$ in the point process framework and to design a bootstrap procedure to calibrate the test.

7.2.3 Goodness-of-fit tests

As stated in Chapter 3, the first-order intensity function can be estimated either parametrically or nonparametrically. The main drawback of parametric techniques is that they can yield unreliable estimates if the assumed model deviates from the true intensity function. For this reason, we need formal tests to assess the goodness-of-fit of parametric models.

Let $\mathbf{X}$ be an inhomogeneous spatial point process with intensity function $\lambda(x)$ observed on a bounded domain $W \in \mathbb{R}^2$, and $\{\lambda_\theta(x); \theta \in \Theta\}$ a class of candidate parametric models. Our aim is to test the null hypothesis: $\mathcal{H}_0 : \lambda(x) = \lambda_{\theta_0}(x)$, for some unknown θ_0. To our knowledge, up to date the only formal goodness-of-fit test for the first-order intensity function is that proposed by Guan (2008a), which uses a L_2 discrepancy measure between the residuals of the fitted model and the number of events observed in W.

The equivalence between the density of event locations and the bivariate density functions suggest developing nonparametric goodness-of-fit test for the first-order intensity analogous to those used to test whether a density function belongs to a given parametric family. For instance, we can consider the L_2 discrepancy introduced by Fan (1994)

$$D_2 = \int \left(\hat{f}_h(x) - f_{\hat{\theta}}(x)\right)^2 dx$$

where $\hat{f}_h(x)$ and $f_{\hat{\theta}}(x)$ are the kernel and parametric estimates of the density function, f, or the L_1 discrepancy introduced by Cao and Lugosi (2005)

$$D_1 = \min_{\theta \in \Theta} \int |\hat{f}_h(x) - f_\theta(x)| dx$$

Assessing the conditions under which these test statistics can be extended to develop goodness-of-fit test for the intensity function of spatial point processes, the analysis of the asymptotic distribution of these tests under the null hypothesis, and the design of bootstrap calibration procedures constitute a new and broad research line.

7.3 Spatio-temporal point processes

Chapter 6 provides a nonparametric separability test for the intensity function of spatio-temporal point processes. To implement this separability test we have used symmetric kernel functions in the spatial ant temporal components of the spatio-temporal intensity, as done by Schoenberg (2004) and Díaz-Avalos et al. (2013)to develop nonparametric separability tests, and by Sarojinie Fernando and Hazelton (2014) to estimate the spatio-temporal relative risk function. However, these approaches may be biased because time has a causal (unidirectional) nature, as stated in the definition of the spatio-temporal conditional intensity (see Section 2.4.1).

Therefore, it seems natural using backward asymmetric kernels in the temporal component to estimate the spatio-temporal intensity function. This one-sided kernels only consider events occurring prior to time t to estimate the intensity function at time t. Future work in this issue may include: (i) analyze the asymptotic properties of the nonparametric intensity estimator with asymmetric kernel in the temporal component; (ii) develop asymmetric edge-correctors for the temporal component of the kernel intensity function; (iii) adapt the current data-driven bandwidth selectors to this new framework; and (iv) check whether using the asymmetric kernel affects the performance of nonparametric separability tests.

USC
UNIVERSIDADE
DE SANTIAGO
DE COMPOSTELA

Appendix A

Supporting information for Chapter 3

A.1 Lack of consistency of the kernel intensity estimator

Let $\{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$ be a realization of a spatial point process in a region $W \subset \mathbb{R}^2$, the kernel estimator of the first-order intensity function, $\lambda(x)$, of $\mathbf{X}$ introduced by Diggle (1985) is given by

$$\hat{\lambda}_h(x) = \frac{1}{p_h(x)} \sum_{i=1}^{N} k_h\left(x - \mathbf{x_i}\right) = \frac{1}{h^2 p_h(x)} \sum_{i=1}^{N} k\left(\left(x - \mathbf{x_i}\right)/h\right)$$

where the kernel function, $k(\cdot)$, is a radially symmetric bivariate probability density function (pdf), $h > 0$ is the smoothing parameter or bandwidth, k_h is the smoothed kernel and $p_h(x) = \int_W h^{-2} k((x-y)/h)dy$ is the edge-correction term.

Considering the infill asymptotic framework in the unbounded domain ($W = \mathbb{R}^2$) and assuming the following regularity conditions:

I1 $\lambda(x)$ is twice differentiable and its first and second-order partial derivatives are continuous and square integrable.

K1 $k(x)$ is a continuous, symmetric, square integrable density function such that $\int_{\mathbb{R}^2} uu^T k(u)du = \mu_2(k) I_2$, with $\mu_2(k) < \infty$.

we can obtain the asymptotic expressions for the bias and variance of $\hat{\lambda}_h(x)$. The expectation of the kernel estimator of $\lambda(x)$ is

$$E[\hat{\lambda}_h(x)] = \frac{1}{h^2}\int_{\mathbb{R}^2} k\left(\frac{x-y}{h}\right)\lambda(y)dy = \int_{\mathbb{R}^2} k(u)\lambda(x-hu)du$$

A second-order Taylor expansion yields

$$E\left[\hat{\lambda}_h(x)\right] = \lambda(x)\int_{\mathbb{R}^2} k(u)du - h\left(\frac{\partial\lambda(x)}{dx_1}\int_{\mathbb{R}^2} u_1k(u)du + \frac{\partial\lambda(x)}{dx_2}\int_{\mathbb{R}^2} u_2k(u)du\right)$$
$$+\frac{h^2}{2}\left(\frac{\partial^2\lambda(x)}{dx_1^2}\int_{\mathbb{R}^2} u_1^2k(u)du + \frac{\partial^2\lambda(x)}{dx_2^2}\int_{\mathbb{R}^2} u_2^2k(u)du + 2\frac{\partial^2\lambda(x)}{dx_1dx_2}\int_{\mathbb{R}^2} u_1u_2k(u)du\right) + o(h^2)$$

By conditions I1 and K1 we obtain the following expression for the bias

$$B\left(\hat{\lambda}_h(x)\right) = E[\hat{\lambda}_h(x)] - \lambda(x) = \frac{h^2}{2}tr\left(D^2\lambda(x)\right)\mu_2(k) + o(h^2)$$

where $D^2\lambda(x)$ is the Hessian matrix of $\lambda(x)$ and for any matrix A, $tr(A)$ denotes the sum of its diagonal entries. Therefore, $B\left(\hat{\lambda}_h(x)\right)$ tends to 0 as $h \to 0$, i.e. $\hat{\lambda}_h(x)$ is asymptotically unbiased.

In virtue of Campbell´s theorem we obtain the following expression for the variance of the kernel intensity estimator

$$Var\left[\hat{\lambda}_h(x)\right] = E\left[\hat{\lambda}_h(x) - E[\hat{\lambda}_h(x)]\right]^2 = \frac{1}{h^4}\int_{\mathbb{R}^2} k\left(\frac{x-y}{h}\right)^2\lambda(y)dy$$
$$+\frac{1}{h^4}\int_{\mathbb{R}^2}\int_{\mathbb{R}^2} k\left(\frac{x-y}{h}\right)k\left(\frac{x-z}{h}\right)\lambda(y)\lambda(z)\left(g(y,z)-1\right)dydz$$

for a Poisson point process $g(y,z) = 1$. Thus, by a Taylor expansion we have

$$Var\left[\hat{\lambda}_h(x)\right] = \frac{1}{h^2}\int_{\mathbb{R}^2} k(u)^2\lambda(x-hu)du = \frac{1}{h^2}\lambda(x)\int_{\mathbb{R}^2} k(u)^2 + o(h^{-2})$$

which does not tend to 0 as $h \to 0$. Then $\hat{\lambda}_h(x)$ is not a consistent estimator of the first-order intensity function.

A.2 Proof of the preliminary results

Let $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$ be a realization of the inhomogeneous spatial Poisson process $\mathbf{X}$ with first-order intensity $\lambda(x)$ observed on a bounded region $W \subset \mathbb{R}^2$, and $N^{-1}\sum_{i=1}^{N} g(\mathbf{x}_i) I\,(N \neq 0)$ for any measurable function g. Given that the number of events, N, has distribution $Poisson\left(\int_W \lambda(x)dx\right) = Poisson\,(m)$, the mean of Z is given by

$$E\,[Z] = E\left[\frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x_i}) I\,(N \neq 0)\right] = E\,[E\,[g(\mathbf{x_i})]\,|N > 0] = \sum_{k=1}^{\infty} E\,[g(\mathbf{x_i})]\,P\,(N = k)$$
$$= \sum_{k=1}^{\infty} \int_W g(x)\lambda_0(x)dx \frac{e^{-m}m^k}{k!} = (1 - e^m)\int_W g(x)\lambda_0(x)dx$$

The variance of Z is $Var(Z) = E\left[Z^2\right] - E\,[Z]^2$ and

$$E\left[Z^2\right] = E\left[\left(\frac{1}{N}\sum_{i=1}^{N} g(\mathbf{x_i}) I\,(N \neq 0)\right)^2\right] = E\left[E\left[\frac{g(\mathbf{x_i})g(\mathbf{x_j})}{N^2}\right]|N > 0\right]$$
$$= \sum_{k=1}^{\infty}\left(\frac{k-1}{k}\int_W\int_W g(x)g(y)\lambda_0(x)\lambda_0(y)dxdy + \frac{1}{k}\int_W g(x)^2\lambda_0(x)dx\right)\frac{e^{-m}m^k}{k!}$$
$$= E\left[\frac{N-1}{N} I\,(N \neq 0)\right]\left(\int_W g(x)\lambda_0(x)dx\right)^2 + E\left[\frac{1}{N} I\,(N \neq 0)\right]\int_W g(x)^2\lambda_0(x)dx$$

Thus, if $A(m) = E\left[\frac{1}{N} I\,(N \neq 0)\right]$ we have

$$Var(Z) = A(m)\int_W g(x)^2\lambda_0(x)dx$$
$$+ \left(E\left[\frac{N-1}{N} I\,(N \neq 0)\right] + (1 - e^{-m})^2\right)\left(\int_W g(x)\lambda_0(x)dx\right)^2$$
$$= A(m)\int_W g(x)^2\lambda_0(x)dx - \left(A(m) - e^{-m} + e^{-2m}\right)\left(\int_W g(x)\lambda_0(x)dx\right)^2$$

A.3 Details about the error measurement of $\hat{\lambda}_{\mathbf{0,H}}(\mathbf{x})$

The kernel estimator for the density of event locations of a spatial point pattern observed on a bounded region W is given by

$$\hat{\lambda}_{0,H}(x) = \frac{\hat{\lambda}_H(x)}{N} I[N \neq 0] = p_H(x)^{-1}|H|^{-1/2}\sum_{i=1}^{N} k\left(H^{-1/2}\,(x - \mathbf{x_i})\right) I[N \neq 0]$$

Considering $g(\mathbf{x_i}) = k_H(x - \mathbf{x_i})$ in (3.2.3) and applying a second-order Taylor expansion, the mean of $\hat{\lambda}_{0,H}(x)$ can be written as follows

$$\begin{aligned}\frac{E\left[\hat{\lambda}_{0,H}(x)\right]}{p_H\left(x\right)\left(1-e^{-m}\right)} &= \int_W k_H\left(x-y\right)\lambda_0(y)dy = \int_{B_{x,H}} \lambda_0\left(x - u^T H^{1/2}\right)k(u)du \\ &= \lambda_0(x)\int_{B_{x,H}} k(u)du - \int_{B_{x,H}} u^T H^{1/2} D\lambda_0(x)k(u)du \\ &+ \frac{1}{2}\int_{B_{x,H}} u^T H D^2\lambda_0(x)uk(u)du + o\left(tr(H)\right)\end{aligned}$$

where $B_{x,H} = \left\{H^{-1}\left(x-y\right); y \in W\right\} = H^{-1}\left(x - W\right)$ and $p_H(x) = \int_W k_H\left(x-y\right)dy = \int_{B_{x,H}} k(u)du$. Thus, the bias of $\hat{\lambda}_{0,H}(x)$ is

$$\begin{aligned}B(x,H) = -e^{-m}\lambda_0(x) &- \frac{\left(1-e^{-m}\right)}{p_H\left(x\right)}\int_{B_{x,H}} u^T H^{1/2} D\lambda_0(x)k(u)du \\ &+ \frac{\left(1-e^{-m}\right)}{p_H\left(x\right)}\left(\frac{1}{2}\int_{B_{x,H}} u^T H D^2\lambda_0(x)uk(u)du + o\left(tr(H)\right)\right)\end{aligned}$$

By conditions I.1-I.3 $B(x,H) \to 0$ when $m \to \infty$ and all terms in $H \to 0$, therefore $\hat{\lambda}_{0,H}(x)$ is asymptotically unbiased. However, as $\lim_{|H|\to 0} B_{x,H}$ depends on W we can not obtain a simpler expression for $B(x,H)$.

When $W = \mathbb{R}^2$, $p_H(x) = 1$ and $B_{x,H} = \mathbb{R}^2$. In addition, by condition I.3, k is symmetric and $\int_{\mathbb{R}^2} uu^T k(u)du = \mu_2(k)I_2$, with $\mu_2(k) < \infty$, thus

$$B(x,H) = -e^{-m}\lambda_0(x) + \frac{\left(1-e^{-m}\right)}{2}\mu_2(k)tr\left(HD^2\lambda_0(x)\right) + o((1-e^{-m})tr(H))$$

has order $O\left(tr(H)\right)$. The same holds at any point x in the interior of W, i.e. at any x such that the support of $k(x)$ is completely contained in $B_{x,H}$. The bias in the boundary of W is $O\left(tr(H)^{1/2}\right)$.

Considering $g(\mathbf{x_i}) = k_H(x - \mathbf{x_i})$ in (3.2.4), the variance of $\hat{\lambda}_{0,H}(x)$ is

$$\begin{aligned}Var(x,H) = &\frac{A(m)}{p_H(x)}\int_W \left(k_H\left(x-y\right)\right)^2\lambda_0(y)dy \\ &- \frac{A(m) - e^{-m} + e^{-2m}}{p_H(x)}\left[\int_W k_H\left(x-y\right)\lambda_0(y)dy\right]^2\end{aligned}$$

By a Taylor expansion, the integral of the first term in the right hand side of this expression is given by

$$\begin{aligned}\int_W \left(k_H\left(x-y\right)\right)^2 \lambda_0(y)dy &= |H|^{-1/2}\int_W k(u)^2\lambda_0(x-u^TH^{1/2})du \\ &= |H|^{-1/2}\lambda_0(x)\int_W k(u)^2du + o(|H|^{-1/2})\end{aligned}$$

Given that $\lambda_0(x)$ is bounded and k is a bivariate density function, the integral in the second term in the right hand side of $Var(x,H)$ has order 1 and

$$Var\left(x,H\right) = \frac{A(m)|H|^{-1/2}}{p_H(x)}\lambda_0(x)\int_{B_{x,H}} k(u)^2du + +o(A(m)|H|^{-1/2})$$

tends to 0 as $m^{-1}|H|^{-1/2}\to 0$ when $m\to\infty$. Thus, the edge effect does not affect the rate of convergence of the variance. When $W=\mathbb{R}^2$ the variance is

$$Var\left(x,H\right) = A(m)|H|^{-1/2}\lambda_0(x)R\left(k\right) + o(A(m)|H|^{-1/2})$$

Therefore if $W=\mathbb{R}^2$ the mean squared error of $\hat{\lambda}_0(x)$ is given by

$$\begin{aligned}MSE(x,H) = B^2(x,H) + Var(x,H) = \frac{\left(1-e^{-m}\right)^2}{4}\mu_2(k)^2tr^2\left(HD^2\lambda_0(x)\right) + e^{-2m}\lambda_0^2(x) \\ -e^{-m}\left(1-e^{-m}\right)\lambda_0(x)\mu_2(k)tr\left(HD^2\lambda_0(x)\right) + A(m)|H|^{-1/2}\lambda_0(x)R(k) \\ +o\left((1-e^{-m})^2tr^2(H)\right) + o\left(e^{-m}(1-e^{-m})tr(H)\right) + o(A(m)|H|^{-1/2})\end{aligned}$$

integrating $MSE(x,H)$ we obtain

$$\begin{aligned}MISE(H) = \frac{\left(1-e^{-m}\right)^2}{4}\mu_2(k)^2\int_{\mathbb{R}^2} tr^2\left(HD^2\lambda_0(x)\right)dx + e^{-2m}\int_{\mathbb{R}^2}\lambda_0^2(x)dx \\ -e^{-m}\left(1-e^{-m}\right)\mu_2(k)\int_{\mathbb{R}^2}\lambda_0(x)tr\left(HD^2\lambda_0(x)\right)dx + A(m)|H|^{-1/2}R(k) \\ +o\left((1-e^{-m})^2tr^2(H)\right) + o\left(e^{-m}(1-e^{-m})tr(H)\right) + o(A(m)|H|^{-1/2})\end{aligned}$$

that yields expression (3.2.7) when $m\to\infty$.

A.4 AMISE for small point patterns

As stated by Marron and Wand (1992) for the kernel density estimator, the remainder term in the MISE can be considerably large for spatial point processes with low mean intensity, m. In order to analyze the effect of m on the discrepancy between the MISE and the AMISE, we need an explicit expression for the MISE. If we look at the bias and variance of $\hat{\lambda}_{0,H}(x)$:.

$$B(x,H) = (1-e^{-m})\frac{(1-e^{-m})}{p_H(x)}\int_W k_H(x-y)\,\lambda_0(y)dy \tag{A.4.1}$$

$$\begin{aligned} Var(x,H) = {} & \frac{A(m)}{p_H(x)}\int_W (k_H(x-y))^2\,\lambda_0(y)dy \\ & - \frac{A(m)-e^{-m}+e^{-2m}}{p_H(x)}\left[\int_W k_H(x-y)\,\lambda_0(y)dy\right]^2 \end{aligned} \tag{A.4.2}$$

the term e^{-m} in the bias of $\hat{\lambda}_{0,H}(x)$ vanishes for relatively small values of m (e.g. $e^{-20} = O(10^{-9})$). In addition if we are dealing with the first-order intensity, $\lambda(x)$, of a spatial point process, which is defined on $\mathbb{R}^2$ and not with a spatial pattern observed in any bounded region, then we do not have any boundary effect. Thus, the bias and variance of $\hat{\lambda}_{0,H}(x)$ are equal to the bias and variance of the bivariate kernel density estimator. Taking this equivalence into account, we have considered spatial point processes with first-order intensity proportional to bivariate normal mixtures, as we know the exact expressions of both MISE and AMISE for the kernel density estimator of bivariate normal mixtures (Chacón and Duong, 2011).

We have simulated inhomogeneous spatial Poisson point processes with the following first-order intensities

$$\lambda_1(x) \propto N\left((0,0), \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}\right)$$

and

$$\lambda_2(x) \propto \frac{3}{4}N\left((-1,-1), \begin{pmatrix} 1 & 0.1 \\ 0.1 & 1 \end{pmatrix}\right) + \frac{1}{4}N\left((1,1), \begin{pmatrix} 1 & -0.9 \\ -0.9 & 1 \end{pmatrix}\right)$$

with different values of m to see the effect of the expected number of events on the accuracy of the AMISE. For each point process and value of m we have computed the full bandwidth matrices minimizing the MISE and the AMISE, we have obtained the

kernel estimator of $\lambda_0(\cdot)$ with both optimal bandwidths and we have computed their MISE and AMISE. We have considered the following discrepancy measures to analyze the goodness-of-fit of AMISE and H_{AMISE} as estimators of MISE and H_{MISE}, respectively.

$$RE_{MISE} = \frac{|AMISE - MISE|}{MISE},\ RE_H = \frac{|MISE\,(H_{AMISE}) - MISE\,(H_{MISE})\,|}{MISE\,(H_{MISE})}$$

RE_{MISE} quantifies the relative weight of the remainder term in the $MISE$, and RE_H the relative error obtained when we use the $AMISE$ instead of the $MISE$ to select the optimal bandwidth. In order to compare the two optimal bandwidth matrices, we have also computed the ratio between their traces, $R_{tr} = tr\,(H_{AMISE})\,/tr\,(H_{MISE})$ and their determinants, $R_{det} = |H_{AMISE}|/|H_{MISE}|$.

		50	100	500	1000	5000	10000
$\lambda_1(x)$	$MISE(H_{MISE})$	0.0065	0.0043	0.0016	0.0011	0.0004	0.0002
	$MISE(H_{AMISE})$	0.0067	0.0044	0.0017	0.0011	0.0004	0.0002
	$AMISE(H_{MISE})$	0.0092	0.0057	0.0019	0.0012	0.0004	0.0003
	$AMISE(H_{AMISE})$	0.0088	0.0056	0.0019	0.0012	0.0004	0.0003
	RE_{MISE}	0.4268	0.3179	0.1684	0.1299	0.0726	0.0569
	RE_H	0.0341	0.0223	0.0081	0.0052	0.0018	0.0012
	R_{tr}	0.8209	0.8542	0.9111	0.9285	0.9574	0.9660
	R_{det}	0.6738	0.7297	0.8302	0.8622	0.9167	0.9331
$\lambda_2(x)$	$MISE(H_{MISE})$	0.0082	0.0057	0.0022	0.0015	0.0005	0.0003
	$MISE(H_{AMISE})$	0.0095	0.0063	0.0023	0.0015	0.0005	0.0003
	$AMISE(H_{MISE})$	0.0186	0.0097	0.0028	0.0017	0.0006	0.0004
	$AMISE(H_{AMISE})$	0.0124	0.0078	0.0027	0.0017	0.0006	0.0004
	RE_{MISE}	1.2610	0.7171	0.2728	0.1949	0.0978	0.0744
	RE_H	0.1584	0.1025	0.0376	0.0243	0.0088	0.0056
	R_{tr}	0.7344	0.7887	0.8739	0.8989	0.9398	0.9518
	R_{det}	0.3758	0.4710	0.6578	0.7203	0.8286	0.8620

TABLE A.1: MISE and AMISE of $\hat{\lambda}_{0,H}(\cdot)$ with optimal bandwidth in terms of MISE (h_{MISE}) and AMISE (H_{AMISE}). Relative weight of the remainder term in the MISE, relative error of the optimal bandwidth in terms of AMISE, and comparison between H_{MISE} and H_{AMISE} for point processes with intensities $\lambda_1(\cdot)$ and $\lambda_2(\cdot)$ and expected intensities ranging between $m = 50$ and $m = 10000$.

Table A.1 shows that the AMISE is always greater than the $MISE$, as shown by Marron and Wand (1992) and Ushakov and Ushakov (2009) for univariate and multivariate kernel density estimation, respectively. In agreement with these authors, we remark that the remainder term can be even larger than the $MISE$ for low mean intensities. We also observe that the convergence of AMISE to MISE is quite slow, notice that for $m = 10000$ $RE_{MISE} > 0.05$ for both simulated point processes. RE_H shows that the discrepancy between $MISE(H_{MISE})$ and $MISE(H_{AMISE})$ is quite small, even when

RE_{MISE} is large, and tends to 0 faster than RE_{MISE}. Thus, we can rely on the AMISE as criterion for bandwidth selection even for point processes with low mean intensity. Finally comparison between H_{MISE} and H_{AMISE} indicates a small bias towards under-smoothing when we use the latter.

A.5 Optimal diagonal bandwidth matrices

If we restrict H to the family of positive-definite diagonal matrices, i.e. $H = diag(h_1^2, h_2^2)$, we obtain a simplified expression for (10). Let $h = (h_1, h_2)^T$ the AMISE of $\hat{\lambda}_{0,h}(x)$ is

$$AMISE(h) = \frac{1}{4}\mu_2(k)^2 \left(h_1^4\psi_{40} + 2h_1^2h_2^2\psi_{22} + h_2^4\psi_{04}\right) + \frac{A(m)}{h_1h_2}R(K)$$

To minimize the AMISE we set to 0 its gradient:

$$DAMISE(h) = \begin{pmatrix} \mu_2(k)^2 \left(h_1^3\psi_{40} + h_1h_2^2\psi_{22}\right) - \dfrac{A(m)R(k)}{h_1^2h_2} \\ \mu_2(k)^2 \left(h_1^2h_2\psi_{22} + h_2^3\psi_{04}\right) - \dfrac{A(m)R(k)}{h_1h_2^2} \end{pmatrix} = 0$$

Solving the equations above we obtain $h_{2,ANISE} = h_{1,AMISE}\left(\psi_{40}/\psi_{04}\right)^{1/4}$ and

$$\mu_2(k)^2 \left(h_1^6 \frac{\psi_{40}^{5/4}\psi_{04}^{1/2}}{\psi_{04}^{3/4}} + h_1^6 \frac{\psi_{40}^{3/4}}{\psi_{04}^{3/4}}\psi_{22}\right) - \frac{A(m)R(k)}{h_1h_2^2} = 0$$

which yields

$$h_{1,AMISE} = \left[\frac{\psi_{04}^{3/4}A(m)R(k)}{\mu_2(k)^2\psi_{40}^{3/4}\left(\psi_{40}^{1/2}\psi_{04}^{1/2} + \psi_{22}\right)}\right]^{1/6}$$

Finally, for $H = h^2I_2$ the AMISE is given by

$$AMISE(h) = \frac{h^4}{4}\mu_2(k)^2\left(\psi_{40} + 2\psi_{22} + \psi_{04}\right) + \frac{A(m)}{h^2}R(K)$$

and the corresponding optimal bandwidth is

$$h_{AMISE} = \left[\frac{2A(m)R(k)}{\mu_2(k)^2\left(\psi_{40} + 2\psi_{22} + \psi_{04}\right)}\right]^{1/6}$$

A.6 Proof of theorem 3.1

Let $\hat{\lambda}_{0,G}(x)$ be the kernel estimator of $\lambda_0(x)$ and $\hat{\lambda}^*_{0,H}(x)$ its bootstrap counterpart. As N^* has distribution $Poisson(\int_W \hat{\lambda}_G(x)dx) = Poisson(\hat{m})$, considering $g(\mathbf{x}^*_\mathbf{i}) = k_H(x - \mathbf{x}^*_\mathbf{i})$ in expression (3.2.3) we obtain

$$\begin{aligned} p_H(x)\, E^*\left[\hat{\lambda}^*_{0,H}(x)\right] &= (1-e^{-\hat{m}})|H|^{-1/2}\int_W k\left(H^{-1/2}(x-y)\right)\hat{\lambda}_{0,G}(y)dy \\ &= (1-e^{-\hat{m}})\int_{B_{x,H}} k(u)\hat{\lambda}_{0,G}\left(x-u^TH^{1/2}\right)du \end{aligned}$$

A second-order Taylor expansion yields the following expression for the bias of $\hat{\lambda}^*_{0,H}(x)$

$$\begin{aligned} B^*(x,H) = &-e^{-\hat{m}}\hat{\lambda}_{0,G}(x) - \frac{(1-e^{-\hat{m}})}{p_H(x)}\int_{B_{x,H}} u^TH^{1/2}D\hat{\lambda}_{0,G}(x)k(u)du \\ &+ \frac{(1-e^{-\hat{m}})}{p_H(x)}\left(\frac{1}{2}\int_{B_{x,H}} u^THD^2\hat{\lambda}_{0,G}(x)uk(u)du + o\left(tr(H)\right)\right) \end{aligned}$$

$\hat{\lambda}^*_{0,H}(x)$ is asymptotically unbiased provided that, by conditions I.1-I.3, $B^*(x,H) \to 0$ when $m \to \infty$ and all terms in $H \to 0$.

If $W = \mathbb{R}^2$, then $p_H(x) = 1$ and $B_{x,H} = \mathbb{R}^2$. In addition, by condition I.3, k is symmetric and $\int_{\mathbb{R}^2} uu^Tk(u)du = \mu_2(k)I_2$, with $\mu_2(k) < \infty$, thus

$$\begin{aligned} B^*(x,H) = &E^*\left[\hat{\lambda}^*_{0,H}(x)\right] - \hat{\lambda}_{0,G}(x) = e^{-\hat{m}}\hat{\lambda}_{0,G}(x) \\ &+ \frac{(1-e^{-\hat{m}})}{2}\mu_2(k)^2tr\left(HD^2\hat{\lambda}_{0,G}(x)\right) + o_p((1-e^{-\hat{m}}))tr(H)) \end{aligned}$$

has order $O_p\left(tr(H)\right)$.

By property (3.2.4), the variance of $\hat{\lambda}^*_{0,H}(x)$ is given by

$$\begin{aligned} Var^*(x,H) = &A(\hat{m})\int_W \left[k_H(x-y)\right]^2\hat{\lambda}_{0,G}(y)dy \\ &- \left(A(\hat{m}) - e^{-\hat{m}} + e^{-2\hat{m}}\right)\left[\int_W k_H(x-y)\,\hat{\lambda}_{0,G}(y)dy\right]^2 \end{aligned}$$

By a Taylor expansion, the first term in the right hand side of the previous expression is given by

$$\begin{aligned}\int_W \left[k_H\left(x-y\right)\right]^2 \hat{\lambda}_{0,G}(y)dy &= |H|^{-1/2}\int_{B_{x,H}} k(u)^2\hat{\lambda}_{0,G}(x-u^TH^{1/2})du \\ &= |H|^{-1/2}\hat{\lambda}_{0,G}(x)\int_{B_{x,H}} k(u)^2du + o_p(|H|^{-1/2})\end{aligned}$$

Given that $\hat{\lambda}_{0,G}(x)$, as a consistent estimator of $\lambda_0(x)$, is bounded and $k(\cdot)$ is a bivariate density function, the integral in the second term in the right hand side of $Var^*(x,H)$ has order 1 and

$$Var^*\left(x,H\right) = \frac{A(\hat{m})|H|^{-1/2}}{p_H(x)}\hat{\lambda}_{0,G}(x)\int_{B_{x,H}} k(u)^2du + o_p\left(A(\hat{m}))|H|^{-1/2}\right)$$

tends to 0 as, by I.1, $m^{-1}|H|^{-1/2}\to 0$ when $m\to\infty$. When $W=\mathbb{R}^2$ the variance is

$$Var^*\left(z,H\right) = A(\hat{m})|H|^{-1/2}\hat{\lambda}_{0,G}(x)R(k) + o_p(A(\hat{m}|H|^{-1/2})$$

Therefore, if we replace $B^*(x,H)$ and $Var^*(x,H)$ in (3.3.2) and let $m\to\infty$ we obtain expression (3.3.4).

A.7 Optimal bandwidths and kernel intensity estimators obtained in the simulation study (Section 3.5)

This section provides detailed results of the simulation study conducted in Section 3.5 to analyze the performance of the plug-in bandwidth selector. Below we outline the optimal bandwidths provided by the different selectors compared for realizations of simulated point processes with $m=500$ and $m=1000$ (Tables A.2- A.9), and the respective kernel intensity estimators for $m=100$ (A.1- A.8).

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.074			0.040	0.45	0.064			0.028	3.12
h2	0.065			0.047	1.19	0.055			0.034	7.33
PLCV	0.113			0.037	1.35	0.098			0.026	1.81
LSCV	0.096			0.036	1.46	0.084			0.026	1.81
Diggle	0.054			0.069	0.01	0.046			0.051	0.03
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.066		0.085	0.037	0.49	0.057		0.074	0.026	3.07
D2	0.059		0.074	0.044	1.2	0.05		0.063	0.032	7.31
D1.s	0.062		0.093	0.035	0.55	0.053		0.081	0.025	3.11
D2.s	0.056		0.082	0.042	1.37	0.047		0.07	0.03	7.3
Scott	0.083		0.101	0.030	< 0.01	0.074		0.09	0.022	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.004	1.4e-5	0.009	0.035	0.59	0.003	8.0e-6	0.007	0.025	3.1
F2.s	0.003	9.2e-6	0.007	0.042	1.39	0.002	5.1e-6	0.005	0.031	7.36
F1.sp	0.004	1.3e-5	0.009	0.035	0.56	0.003	1.2e-5	0.007	0.025	3.2
F2.sp	0.003	8.3e-6	0.007	0.042	1.4	0.002	8.6e-6	0.005	0.031	7.63
Flscv	0.001	-2.2e-5	0.007	0.095	4.64	0.001	-3.7e-5	0.005	0.067	6.68

TABLE A.2: Comparison of bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_1(x)$. Mean of the optimal bandwidths provided by each criteria, mean of the ISEs (MISE) of the respective kernel estimator for the 100 realizations of the point process and computational cost of each bandwidth selector. Labels "1" and "2" indicate 1 and 2-stage plug-in bandwidth selection; h: scalar bandwidth ($H = h^2 I_2$); D: Diagonal matrix ($H = diag(h_1^2, h_2^2)$); F: full matrix. "s" indicates pre-scaled and "sp" pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.072			0.093	0.58	0.063			0.050	4.16
h2	0.064			0.105	1.37	0.055			0.057	9.75
PLCV	0.091			0.081	1.5	0.087			0.045	2.03
LSCV	0.089			0.086	1.56	0.080			0.094	2.01
Diggle	0.054			0.130	< 0.01	0.047			0.075	0.02
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.066		0.082	0.090	0.53	0.056		0.073	0.047	4.11
D2	0.059		0.072	0.102	1.36	0.049		0.063	0.055	9.73
D1.s	0.062		0.090	0.087	0.53	0.053		0.080	0.046	4.23
D2.s	0.055		0.080	0.098	1.33	0.047		0.069	0.053	11.05
Scott	0.082		0.102	0.073	< 0.01	0.074		0.090	0.039	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.004	-3.0e-7	0.008	0.087	0.55	0.003	-2.5e-6	0.006	0.046	5.15
F2.s	0.003	1.9e-6	0.006	0.098	1.37	0.002	-4.1e-6	0.005	0.053	10.61
f1.sp	0.004	5.7e-5	0.008	0.087	0.55	0.003	-1.2e-8	0.006	0.046	4.35
F2.sp	0.003	5.0e-5	0.006	0.098	1.35	0.002	-2.4e-6	0.005	0.053	10.03
Flscv	0.001	-2.3e-5	0.007	0.162	1.34	0.001	1.9e-5	0.005	0.043	10.1

TABLE A.3: Comparison of bandwidth selection procedures for Thomas cluster point processes with intensity $\lambda_1(x)$. See details in the caption of Table A.2 for complete information.

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.050			0.102	1	0.041			0.067	5.51
h2	0.047			0.105	2.56	0.040			0.068	12.42
PLCV	0.067			0.132	2	0.056			0.085	2.89
LSCV	0.051			0.106	2.26	0.046			0.067	3.66
Diggle	0.049			0.118	0.03	0.050			0.077	0.06
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.045		0.056	0.092	0.97	0.038		0.045	0.062	5.44
D2	0.043		0.051	0.099	2.49	0.037		0.043	0.065	12.88
D1.s	0.040		0.083	0.070	0.97	0.034		0.070	0.046	5.36
D2.s	0.039		0.079	0.073	2.47	0.033		0.069	0.047	12.52
Scott	0.048		0.102	0.071	< 0.01	0.039		0.090	0.044	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.002	-0.001	0.007	0.066	1.01	0.001	-0.001	0.005	0.044	5.44
F2.s	0.002	-0.001	0.006	0.069	2.53	0.001	-0.0005	0.005	0.045	13.01
f1.sp	0.002	-0.001	0.008	0.062	1	0.001	-0.001	0.006	0.041	5.45
F2.sp	0.002	-0.001	0.007	0.064	2.5	0.001	-0.001	0.006	0.041	12.28
Flscv	0.002	-0.001	0.003	0.121	3.21	0.002	-0.0004	0.002	0.076	23.47

TABLE A.4: Comparison of bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_2(x)$. See details in the caption of Table A.2 for complete information.

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.064			0.246	0.75	0.05			0.176	4.2
h2	0.056			0.264	1.89	0.045			0.181	9.93
PLCV	0.049			0.295	1.61	0.046			0.165	2.05
LSCV	0.048			0.3	1.62	0.044			0.175	2.15
Diggle	0.046			0.319	0.01	0.047			0.147	0.03
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.059		0.072	0.232	0.71	0.046		0.056	0.144	4.18
D2	0.053		0.06	0.258	1.88	0.042		0.048	0.153	9.73
D1.s	0.055		0.083	0.216	0.75	0.042		0.071	0.141	4.12
D2.s	0.05		0.071	0.238	1.93	0.039		0.063	0.148	10.98
Scott	0.076		0.102	0.213	< 0.01	0.055		0.091	0.155	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.003	-0.001	0.007	0.213	0.78	0.002	-0.0005	0.005	0.185	5.16
F2.s	0.003	-0.0003	0.005	0.237	1.86	0.002	-0.0003	0.004	0.21	10.31
F1.sp	0.003	-0.001	0.007	0.21	0.75	0.002	-0.001	0.005	0.19	4.83
F2.sp	0.003	-0.001	0.005	0.232	1.93	0.002	-0.001	0.004	0.15	10.09
Flscv	0.002	-0.0002	0.002	0.351	3.09	0.002	-0.0001	0.001	0.185	16.75

TABLE A.5: Comparison of bandwidth selectors for Thomas cluster point processes with intensity $\lambda_2(x)$. See details in the caption of Table A.2 for complete information.

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.028			2.363	0.75	0.020			1.194	4.64
h2	0.021			1.488	1.95	0.015			0.820	11.01
PLCV	0.033			3.075	1.92	0.026			2.094	2.79
LSCV	0.015			1.213	2.19	0.013			0.764	3.52
Diggle	0.014			1.233	0.03	0.013			0.787	0.04
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.022		0.045	1.618	0.75	0.015		0.032	0.749	4.54
D2	0.017		0.031	1.028	1.92	0.012		0.022	0.575	10.72
D1.s	0.018		0.067	1.379	0.75	0.012		0.050	0.653	4.79
D2.s	0.013		0.050	0.843	1.94	0.010		0.039	0.458	10.69
Scott	0.038		0.102	4.179	< 0.01	0.028		0.091	2.952	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.000	-0.001	0.006	0.896	0.78	0.000	-0.001	0.004	0.369	4.66
F2.s	0.000	-0.0005	0.003	0.574	1.97	0.000	-0.0003	0.002	0.302	11.07
F1.sp	0.001	-0.002	0.008	0.661	0.78	0.000	-0.001	0.006	0.245	4.59
F2.sp	0.000	-0.001	0.004	0.449	1.98	0.000	-0.001	0.003	0.228	10.80
Flscv	0.000	-0.001	0.003	0.584	3.20	0.000	-0.0004	0.002	0.343	16.11

TABLE A.6: Comparison of bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_3(x)$. See details in the caption of Table A.2 for complete information.

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.088			0.101	0.49	0.076			0.086	2.81
h2	0.081			0.101	1.22	0.069			0.085	6.51
PLCV	0.115			0.106	1.35	0.070			0.088	1.64
LSCV	0.102			0.105	1.41	0.068			0.088	1.7
Diggle	0.056			0.118	< 0.01	0.052			0.091	0.02
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.084		0.093	0.101	0.45	0.072		0.080	0.086	2.72
D2	0.078		0.085	0.101	1.19	0.066		0.072	0.085	6.52
D1.s	0.086		0.092	0.101	0.48	0.074		0.079	0.086	0.52
D2.s	0.079		0.084	0.101	1.18	0.067		0.071	0.085	1.21
Scott	0.105		0.100	0.101	< 0.01	0.094		0.089	0.089	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.007	-1.9e-4	0.008	0.101	0.49	0.005	-1.1e-4	0.006	0.086	0.49
F2.s	0.006	-9.6e-5	0.007	0.101	1.2	0.004	-2.4-5	0.005	0.086	1.21
F1.sp	0.007	-2.4e-4	0.008	0.101	0.49	0.005	-1.2e-4	0.006	0.086	0.48
F2.sp	0.006	-1.4e-4	0.007	0.101	1.22	0.004	-3.2e-5	0.005	0.086	1.22
Flscv	0.003	2.4e-4	0.005	0.126	1.71	0.002	9.0e-5	0.003	0.101	11.61

TABLE A.7: Comparison of bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_4(x)$. See details in the caption of Table A.2 for complete information.

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.075			0.149	0.61	0.062			0.112	3.24
h2	0.066			0.128	1.49	0.054			0.091	7.66
PLCV	0.053			0.116	1.63	0.040			0.085	1.94
LSCV	0.052			0.113	1.79	0.042			0.080	2.02
Diggle	0.049			0.122	0.01	0.044			0.084	0.03
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.075		0.075	0.149	0.57	0.062		0.062	0.112	3.45
D2	0.067		0.066	0.128	1.59	0.054		0.054	0.091	7.92
D1.s	0.075		0.075	0.149	0.6	0.062		0.062	0.112	3.23
D2.s	0.067		0.066	0.128	1.6	0.054		0.054	0.091	7.82
Scott	0.101		0.101	0.233	< 0.01	0.090		0.090	0.218	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.006	-2.5e-5	0.006	0.149	0.62	0.004	5.0e-6	0.004	0.113	3.62
F2.s	0.004	-1.8e-5	0.004	0.128	1.52	0.003	3.6e-6	0.003	0.091	7.82
f1.sp	0.006	-1.9e-6	0.006	0.150	0.58	0.004	-8.3e-6	0.004	0.113	3.25
F2.sp	0.004	-1.0e-6	0.004	0.128	1.61	0.003	-4.4e-6	0.003	0.091	8.32
Flscv	0.002	-2.7e-5	0.002	0.127	4.99	0.001	-3.6e-5	0.002	0.087	8.93

TABLE A.8: Comparison of bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_5(x)$. See details in the caption of Table A.2 for complete information.

	$m = 500$					$m = 1000$				
	h			MISE	cost	h			MISE	cost
h1	0.077			0.109	0.63	0.065			0.072	3.76
h2	0.069			0.099	1.61	0.057			0.063	9.21
PLCV	0.058			0.096	1.58	0.050			0.063	2.07
LSCV	0.059			0.095	1.59	0.049			0.061	2.09
Diggle	0.053			0.098	0.03	0.049			0.064	0.03
	h_1		h_2	$MISE$	cost	h_1		h_2	$MISE$	cost
D1	0.077		0.076	0.109	0.63	0.065		0.065	0.072	4.1
D2	0.069		0.068	0.099	1.57	0.058		0.057	0.063	10.59
D1.s	0.077		0.076	0.109	0.62	0.065		0.065	0.072	4.78
D2.s	0.069		0.068	0.099	1.7	0.058		0.057	0.064	10.77
Scott	0.097		0.097	0.149	< 0.01	0.086		0.087	0.113	< 0.01
	H_{11}	H_{21}	H_{22}	$MISE$	cost	H_{11}	H_{21}	H_{22}	$MISE$	cost
F1.s	0.006	-0.002	0.006	0.095	0.66	0.005	-0.001	0.005	0.061	4.25
F2.s	0.005	-0.001	0.005	0.087	1.66	0.004	-0.001	0.004	0.054	9.95
F1.sp	0.007	-0.002	0.007	0.091	0.65	0.005	-0.002	0.005	0.058	4.54
F2.sp	0.005	-0.002	0.005	0.084	1.72	0.004	-0.002	0.004	0.052	9.72
Flscv	0.004	-0.002	0.004	0.090	2.76	0.003	-0.001	0.002	0.058	11.46

TABLE A.9: Comparison of bandwidth selectors for inhomogeneous Poisson point processes with intensity $\lambda_6(x)$. See details in the caption of Table A.2 for complete information.

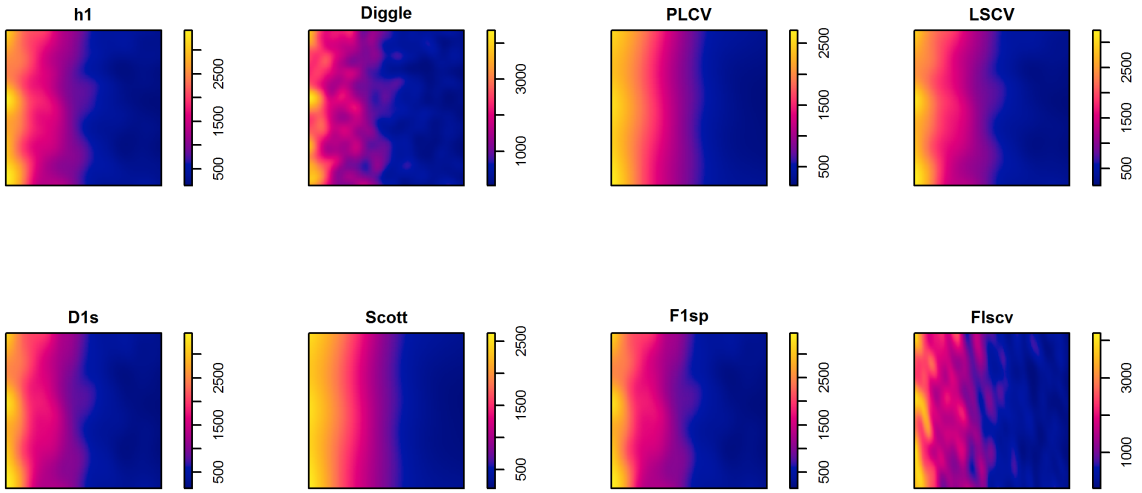


FIGURE A.1: Kernel intensity estimators for an inhomogeneous Poisson point process with first-order intensity $\lambda_1(x)$ and $m = 1000$. The number "1" indicates 1-stage plug-in bandwidth selection. h1: scalar bandwidth parameter; D1s: Diagonal matrix with pre-scaled data; F1sp: Full matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

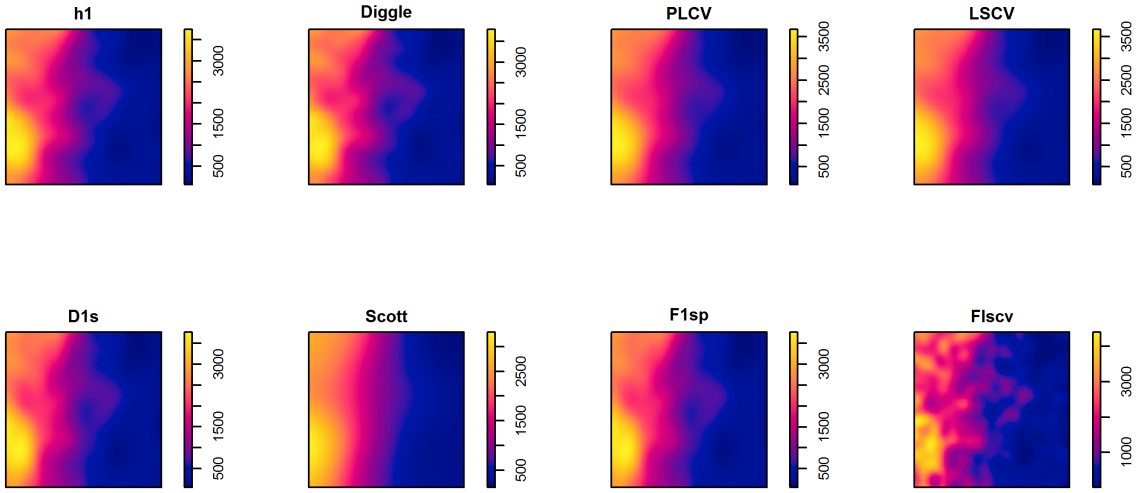


FIGURE A.2: Kernel intensity estimation for a Thomas Cluster point process with first-order intensity $\lambda_1(x)$ and $m = 1000$. The label "1" indicates 1-stage plug-in bandwidth selection. h1: scalar bandwidth parameter; D1s: Diagonal matrix with pre-scaled data; F1sp: Full matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

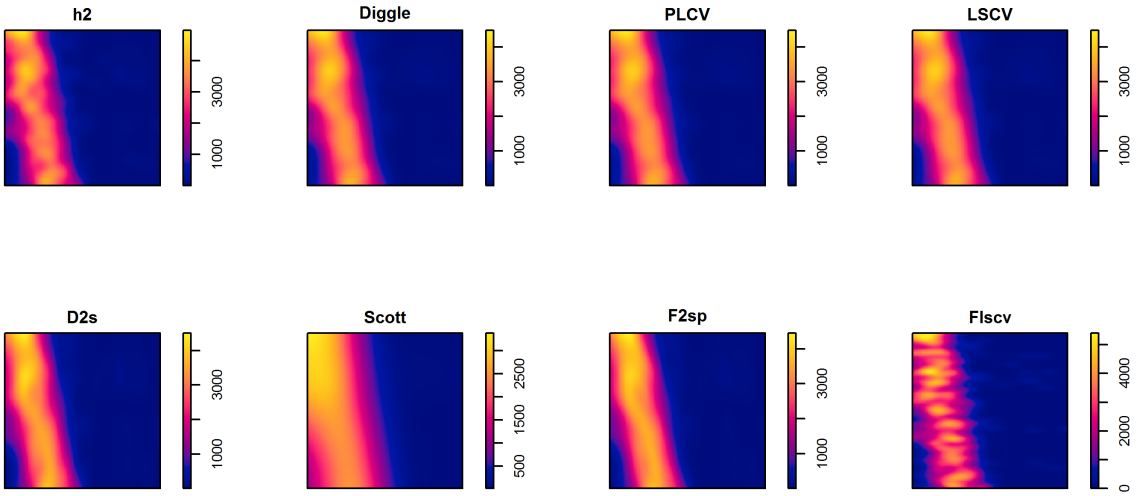


FIGURE A.3: Kernel intensity estimation for an inhomogeneous Poisson point process with first-order intensity $\lambda_2(x)$ and $m = 1000$. The label "2" indicates 1-stage plug-in bandwidth selection. h2: scalar bandwidth parameter; D2s: Diagonal matrix with pre-scaled data; F2sp: Full matrix with pre-sphered data;PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

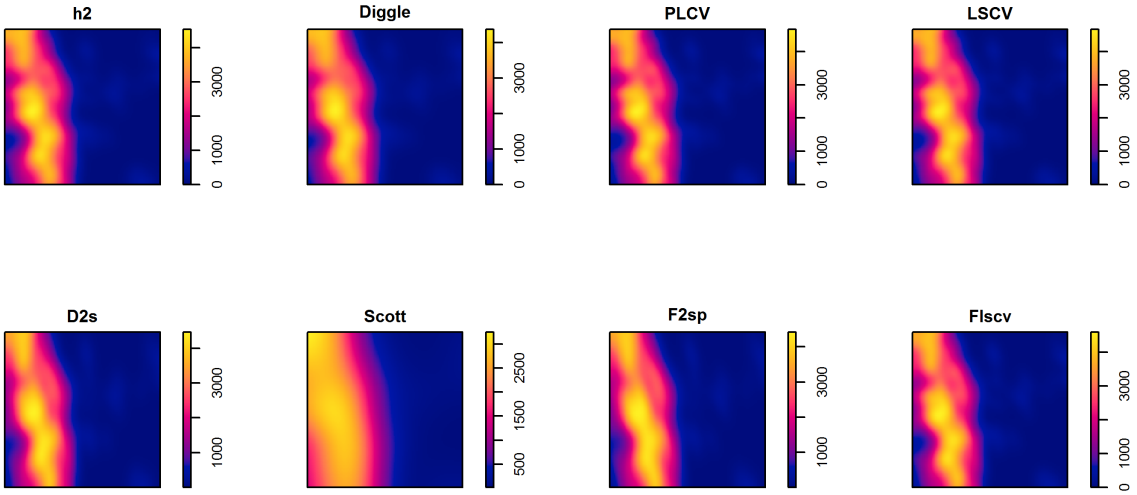


FIGURE A.4: Kernel intensity estimation for a Thomas Cluster point process with first-order intensity $\lambda_2(x)$ and $m = 1000$. The label "1" indicates 2-stage plug-in bandwidth selection. h2: scalar bandwidth parameter; D2s: Diagonal matrix with pre-scaled data; F2sp: Full matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

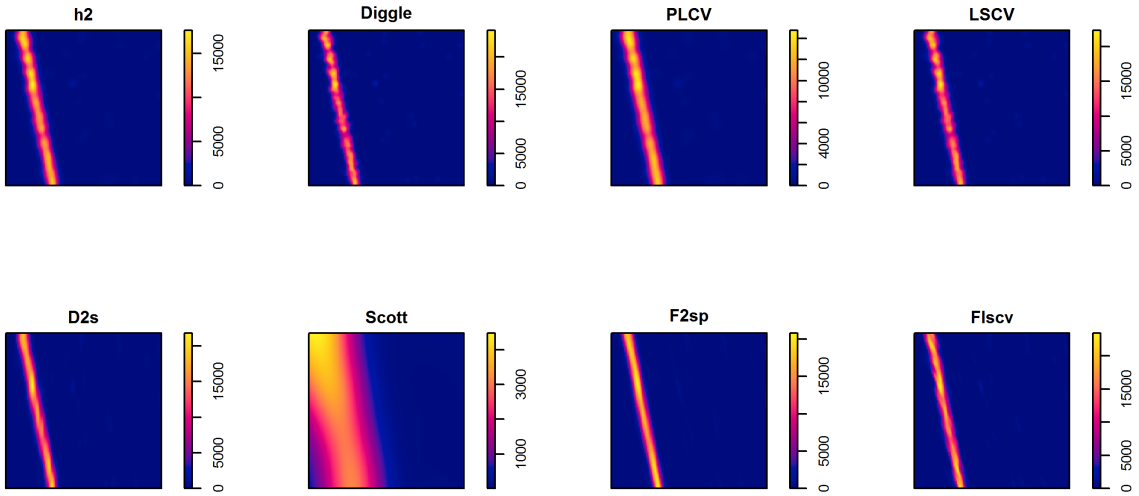


FIGURE A.5: Kernel intensity estimation for an inhomogeneous Poisson point process with first-order intensity $\lambda_3(x)$ and $m = 1000$. The label "2" indicates 1-stage plug-in bandwidth selection. h2: scalar bandwidth parameter; D2s: Diagonal matrix with pre-scaled data; F2sp: Full matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

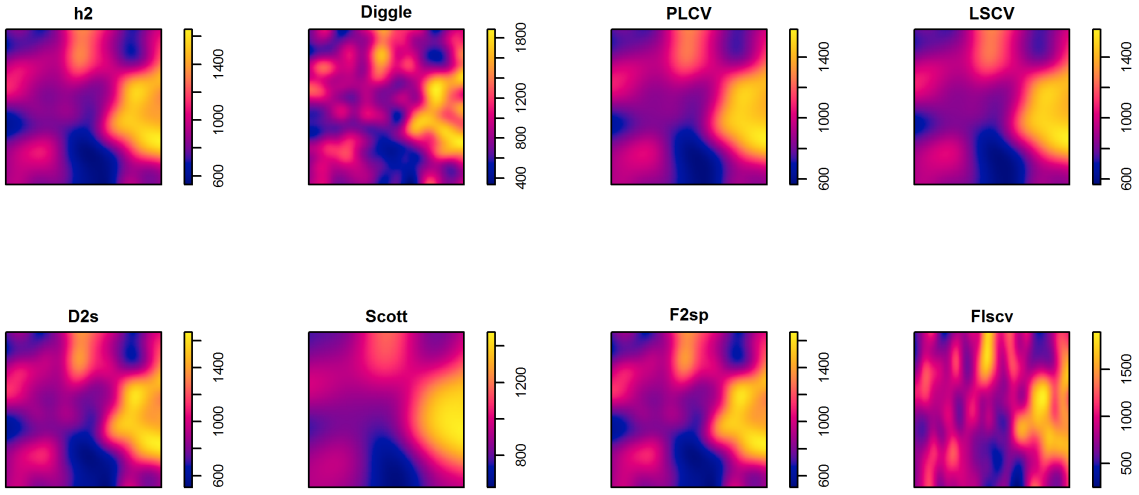


FIGURE A.6: Kernel intensity estimation for an inhomogeneous Poisson point process with first-order intensity $\lambda_1(x)$ and $m = 1000$. The label "1" indicates 1-stage plug-in bandwidth selection; h1: scalar bandwidth parameter; D1s: Diagonal matrix with pre-scaled data; F1sp: Full matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

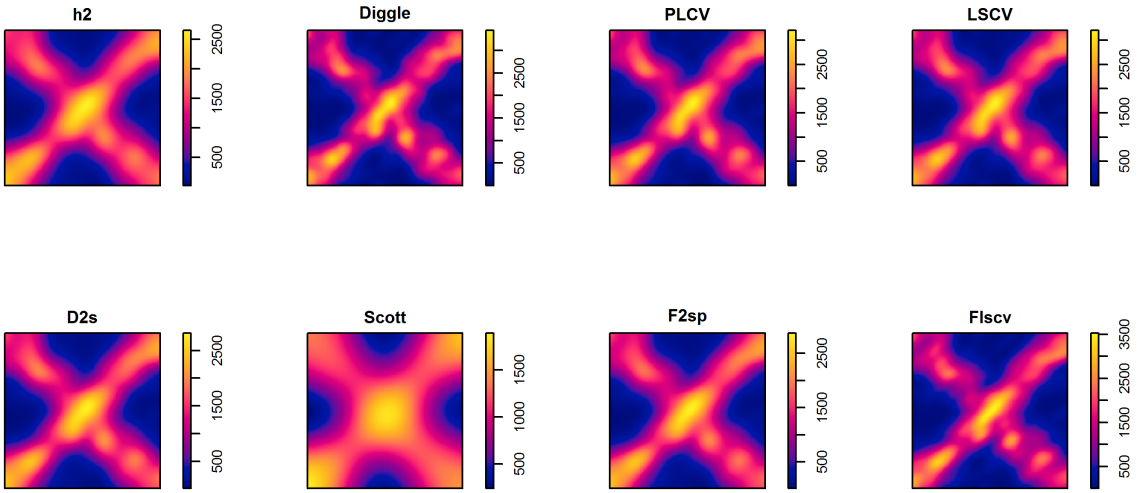


FIGURE A.7: Kernel intensity estimation for an inhomogeneous Poisson poison process with first-order intensity $\lambda_5(x)$ and $m = 1000$. The label "2" indicates 1-stage plug-in bandwidth selection. h2: scalar bandwidth parameter; D2s: Diagonal matrix with pre-scaled data; F2sp: Full matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

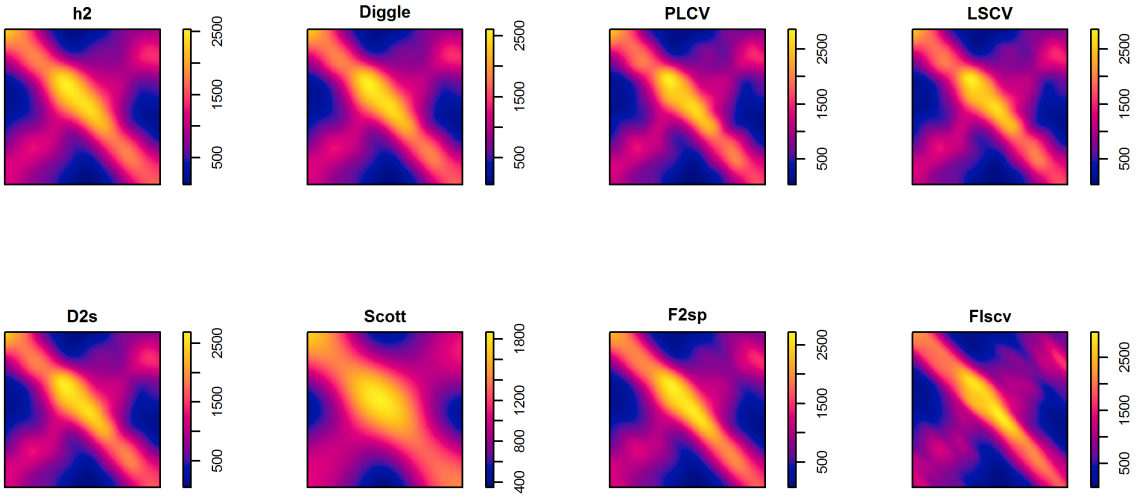


FIGURE A.8: Kernel intensity estimation for an inhomogeneous Poisson point process with first-order intensity $\lambda_6(x)$ and $m = 1000$. The label "2" indicates 1-stage plug-in bandwidth selection. h2: scalar bandwidth parameter; D2s: Diagonal matrix with pre-scaled data; F2sp: Full matrix with pre-sphered data; PLCV: pseudolikelihood cross-validation; LSCV: least-squares cross-validation; Flscv: least-squares cross-validation matrix; Diggle: Diggle´s selector; Scott: Scott´s rule of thumbs.

Appendix B

Supporting information for Chapter 5

In this Appendix we provide graphical support for the analysis of the spatial patterns of wildfires in Galicia during 1999-2008. Section B.1 shows the kernel intensity functions for the spatial pattern of wildfires classified by size and cause for the ten years under study. Section B.2 shows the inhomogeneous L-test for the spatial patterns of wildfires by size and cause for each year, which results were outlined in table 5.6.

B.1 First-order intensity functions of wildfires

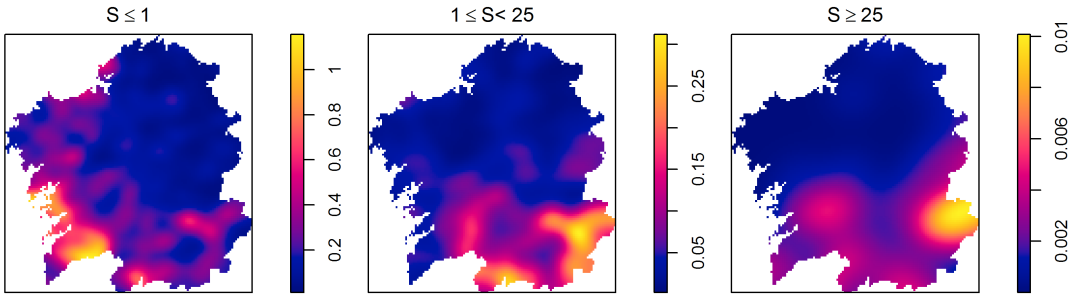


FIGURE B.1: Kernel intensity estimator of the wildfires registered in 1999 by burned area (different scales)

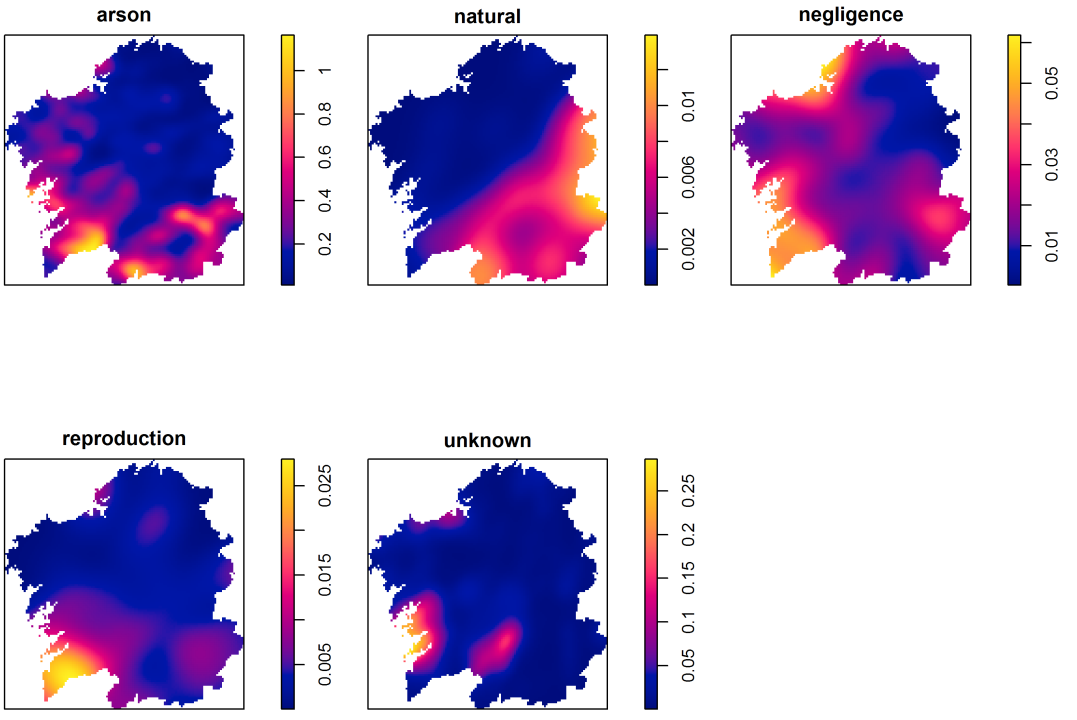


FIGURE B.2: Kernel intensity estimator of the wildfires registered in 1999 by cause (different scales)

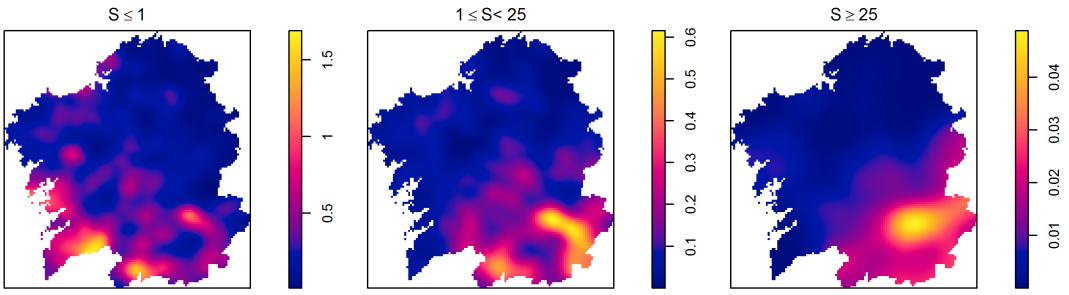


FIGURE B.3: Kernel intensity estimator of the wildfires registered in 2000 by burned area (different scales)

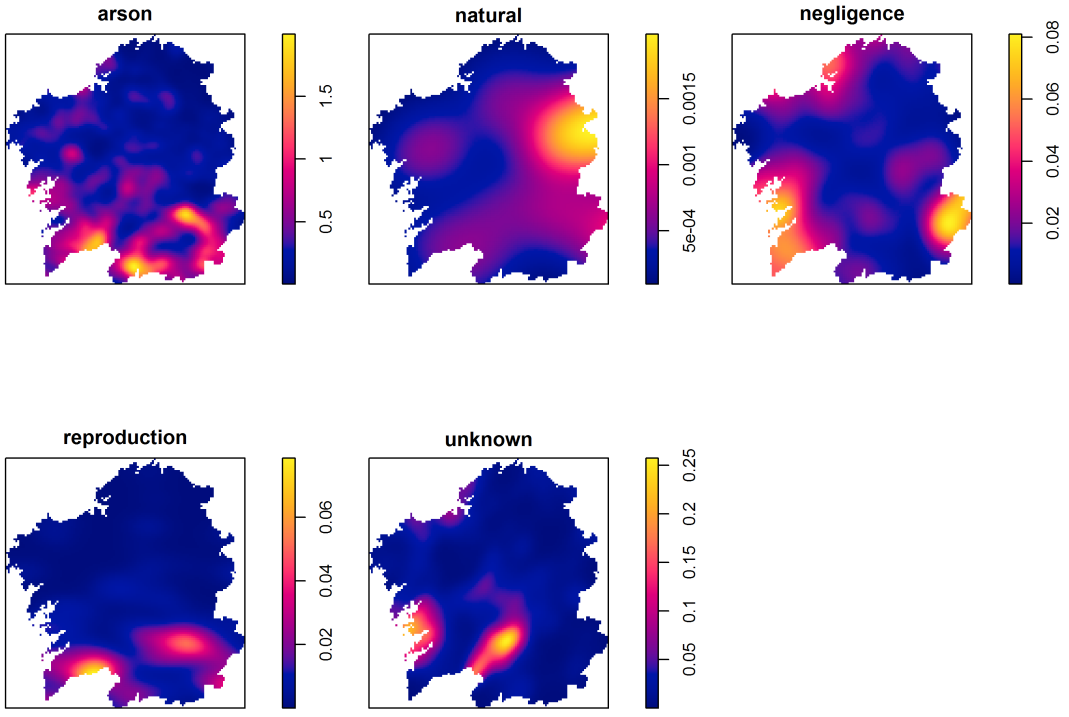


FIGURE B.4: Kernel intensity estimator of the wildfires registered in 2000 by cause (different scales)

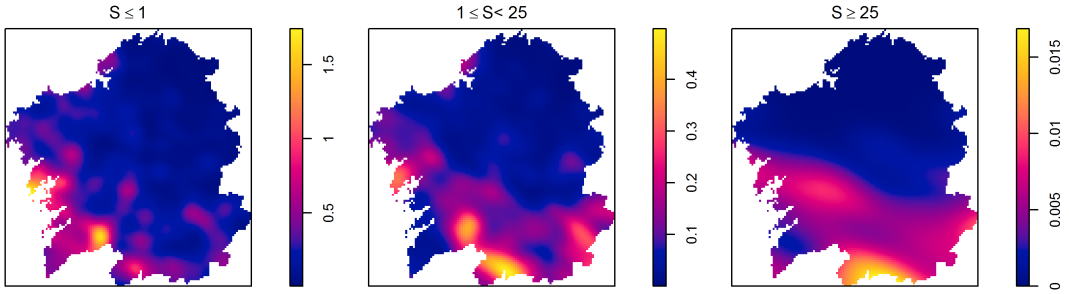


FIGURE B.5: Kernel intensity estimator of the wildfires registered in 2001 by burned area (different scales)

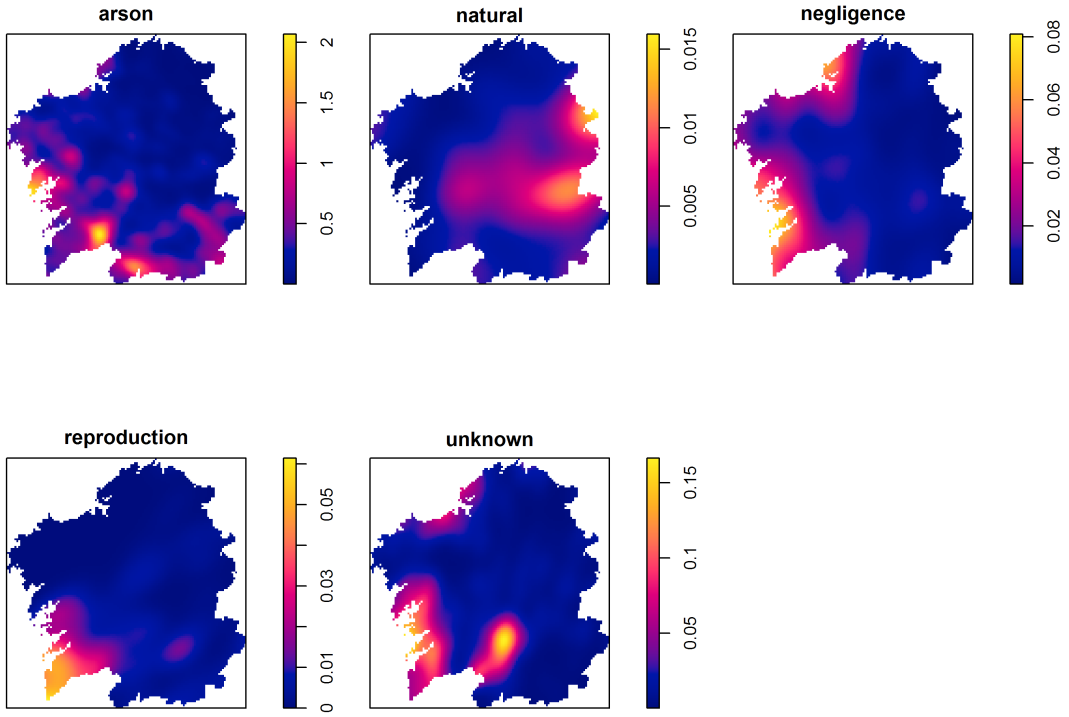


FIGURE B.6: Kernel intensity estimator of the wildfires registered in 2001 by cause (different scales)

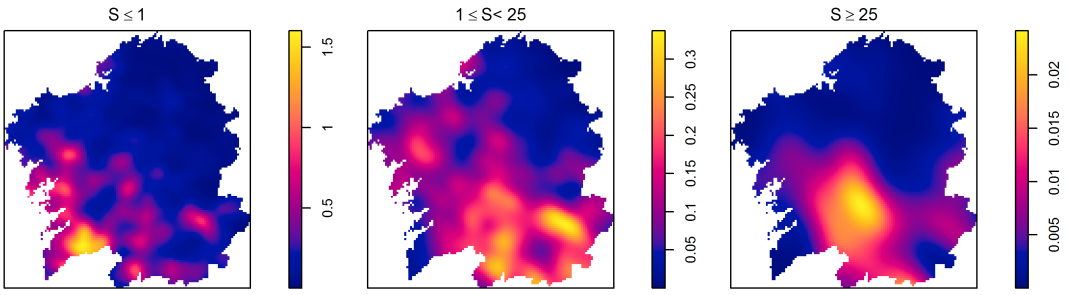


FIGURE B.7: Kernel intensity estimator of the wildfires registered in 2002 by burned area (different scales)

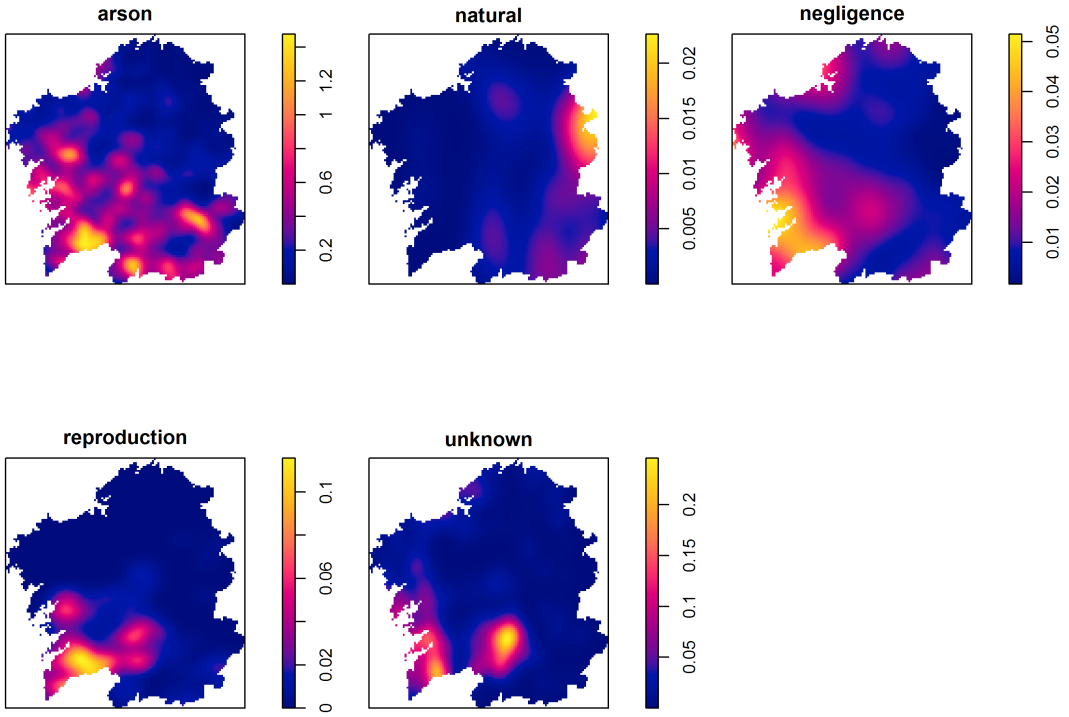


FIGURE B.8: Kernel intensity estimator of the wildfires registered in 2002 by cause (different scales)

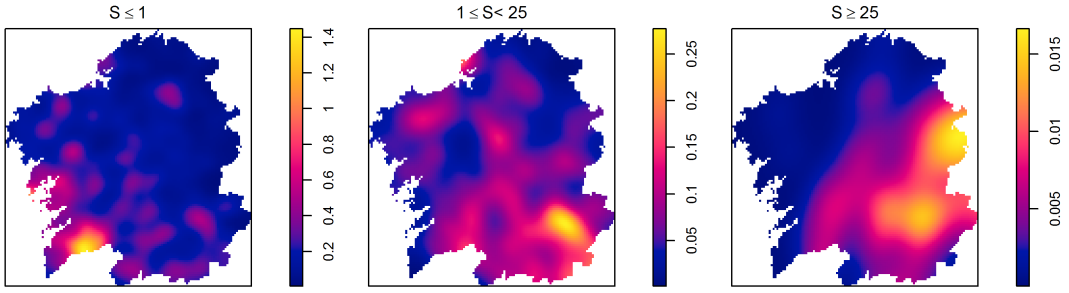


FIGURE B.9: Kernel intensity estimator of the wildfires registered in 2003 by burned area (different scales)

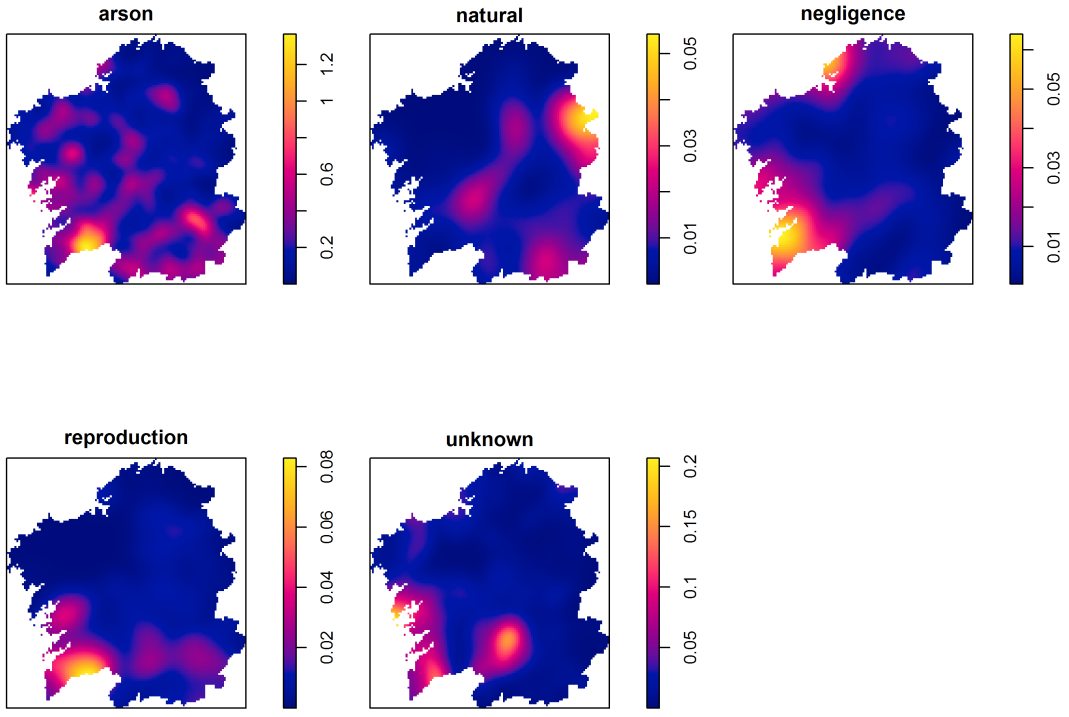


FIGURE B.10: Kernel intensity estimator of the wildfires registered in 2003 by cause (different scales)

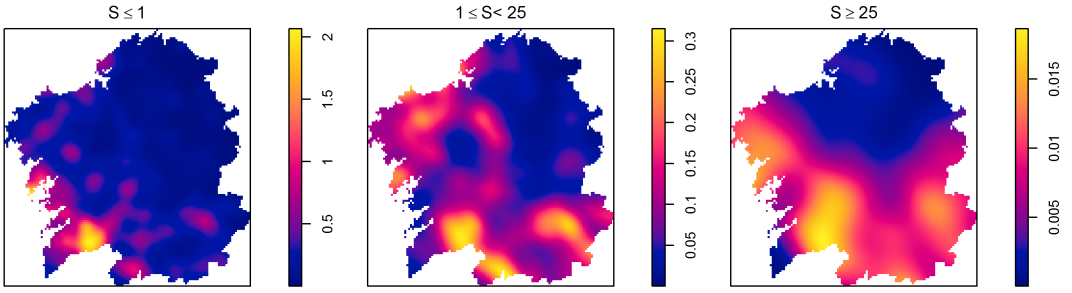


FIGURE B.11: Kernel intensity estimator of the wildfires registered in 2004 by burned area (different scales)

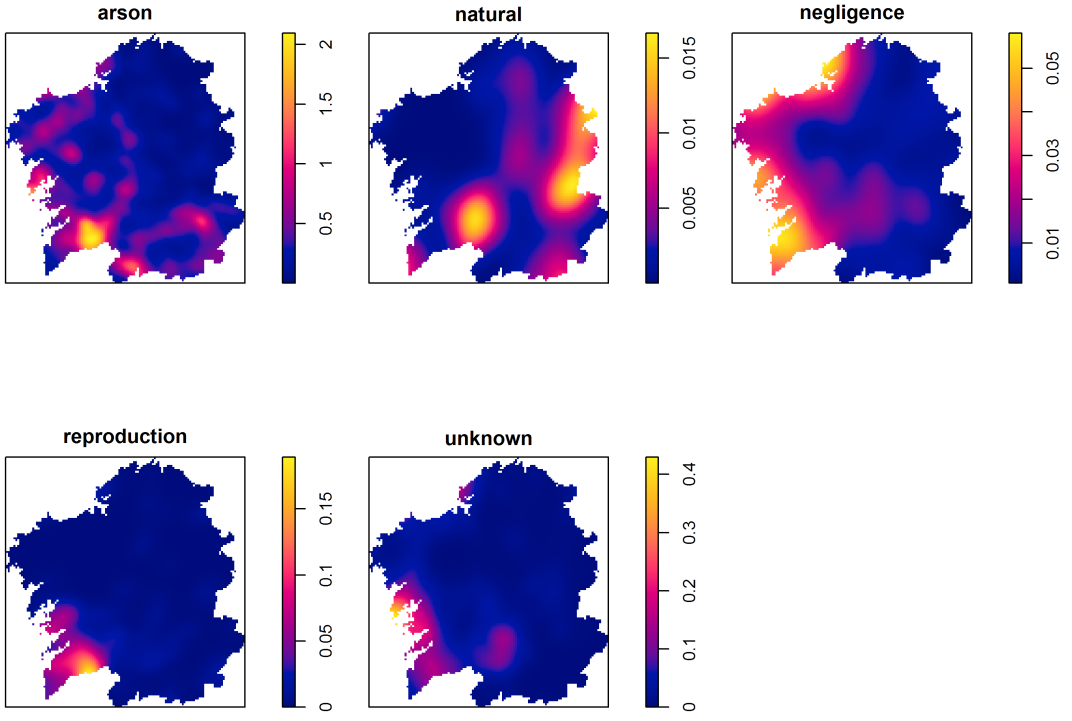


FIGURE B.12: Kernel intensity estimator of the wildfires registered in 2004 by cause (different scales)

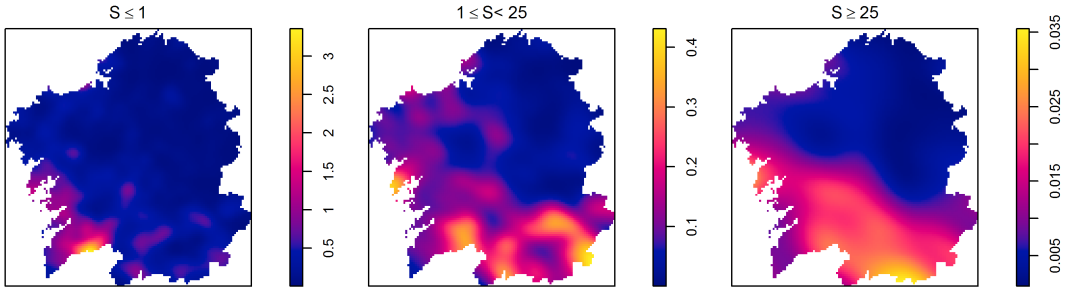


FIGURE B.13: Kernel intensity estimator of the wildfires registered in 2005 by burned area (different scales)

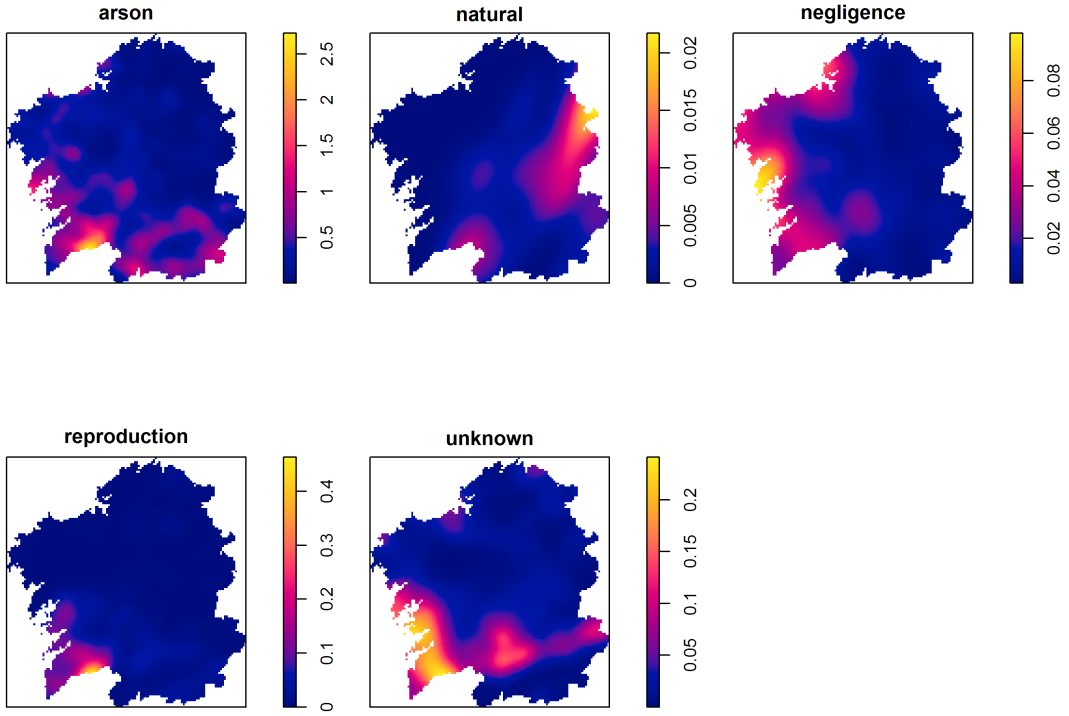


FIGURE B.14: Kernel intensity estimator of the wildfires registered in 2005 by cause (different scales)

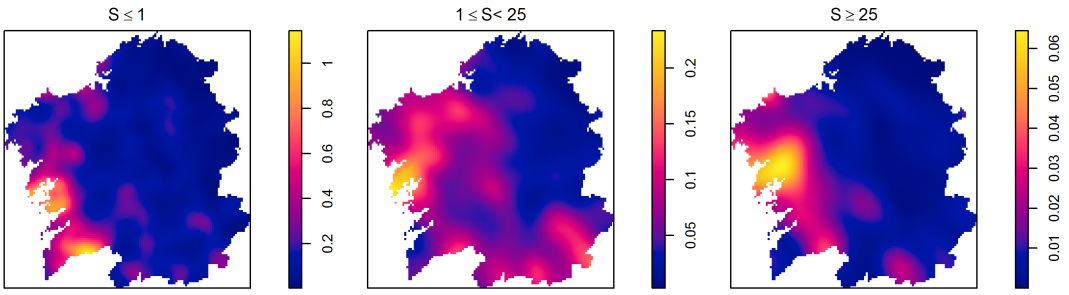


FIGURE B.15: Kernel intensity estimator of the wildfires registered in 2006 by burned area (different scales)

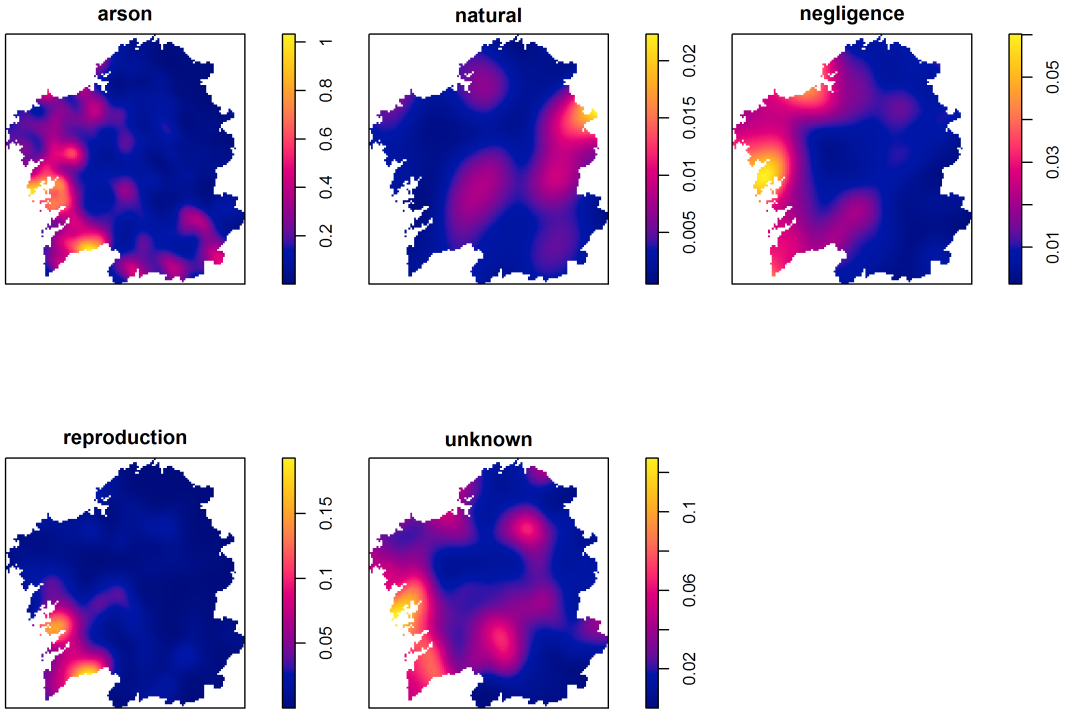


FIGURE B.16: Kernel intensity estimator of the wildfires registered in 2006 by cause (different scales)

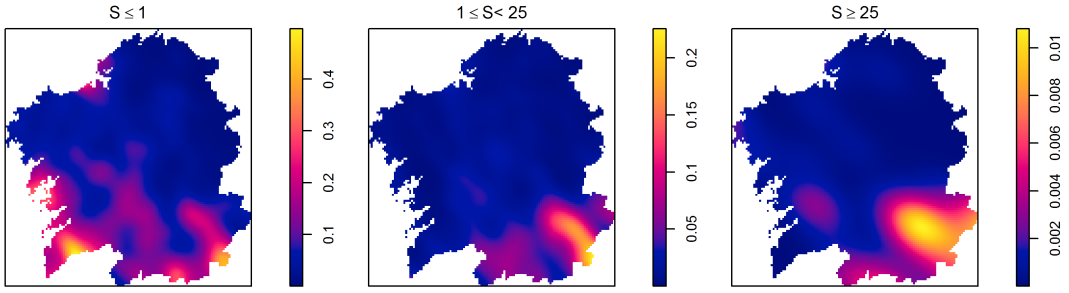


FIGURE B.17: Kernel intensity estimator of the wildfires registered in 2007 by burned area (different scales)

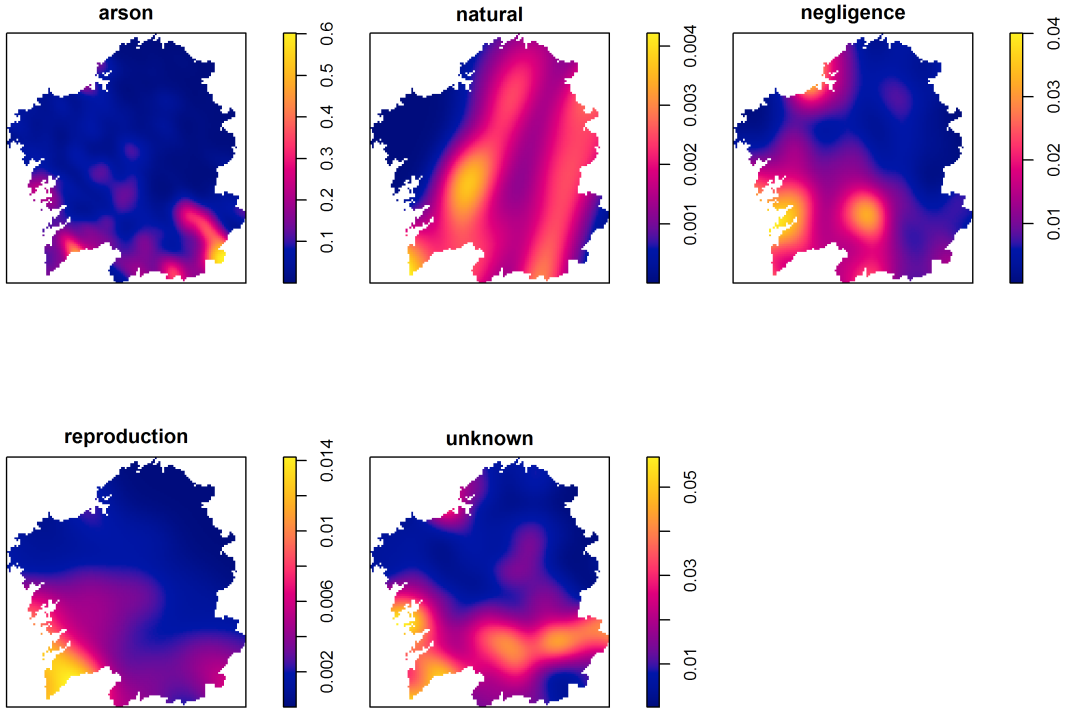


FIGURE B.18: Kernel intensity estimator of the wildfires registered in 2007 by cause (different scales)

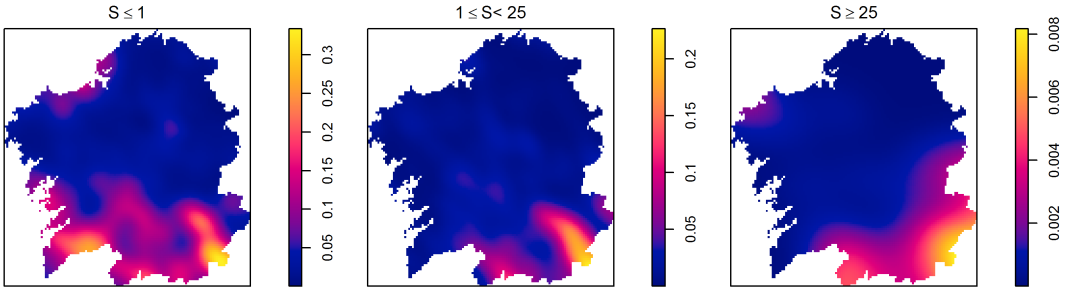


FIGURE B.19: Kernel intensity estimator of the wildfires registered in 2008 by burned area (different scales)

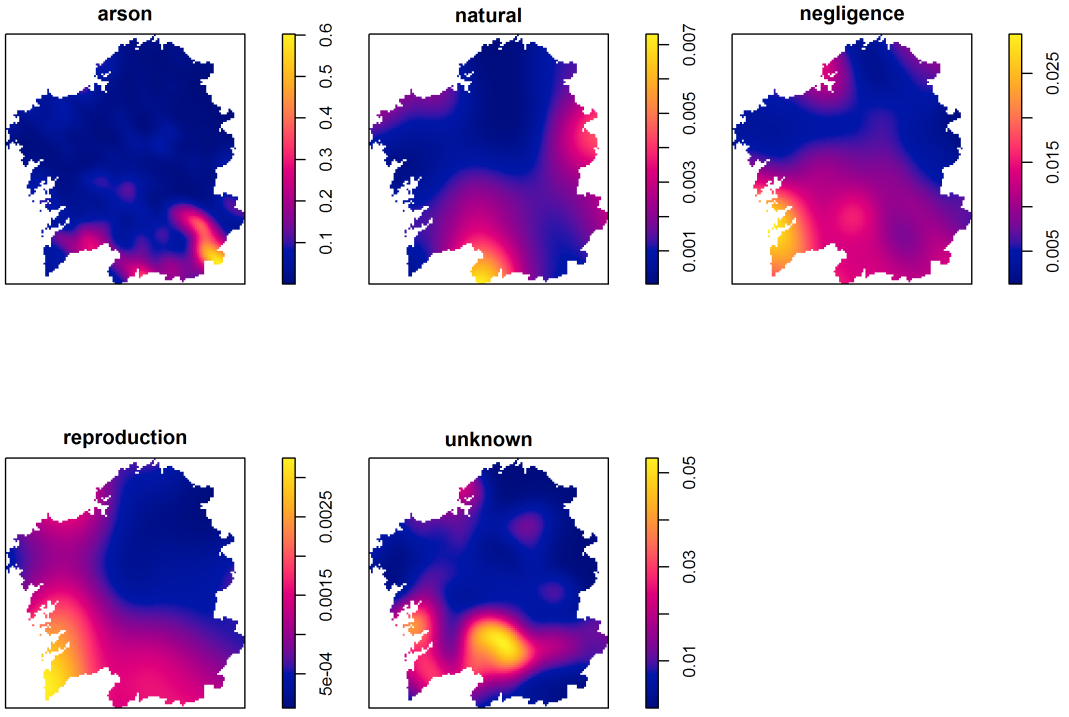


FIGURE B.20: Kernel intensity estimator of the wildfires registered in 2008 by cause (different scales)

B.2 Inhomogeneous L-tests

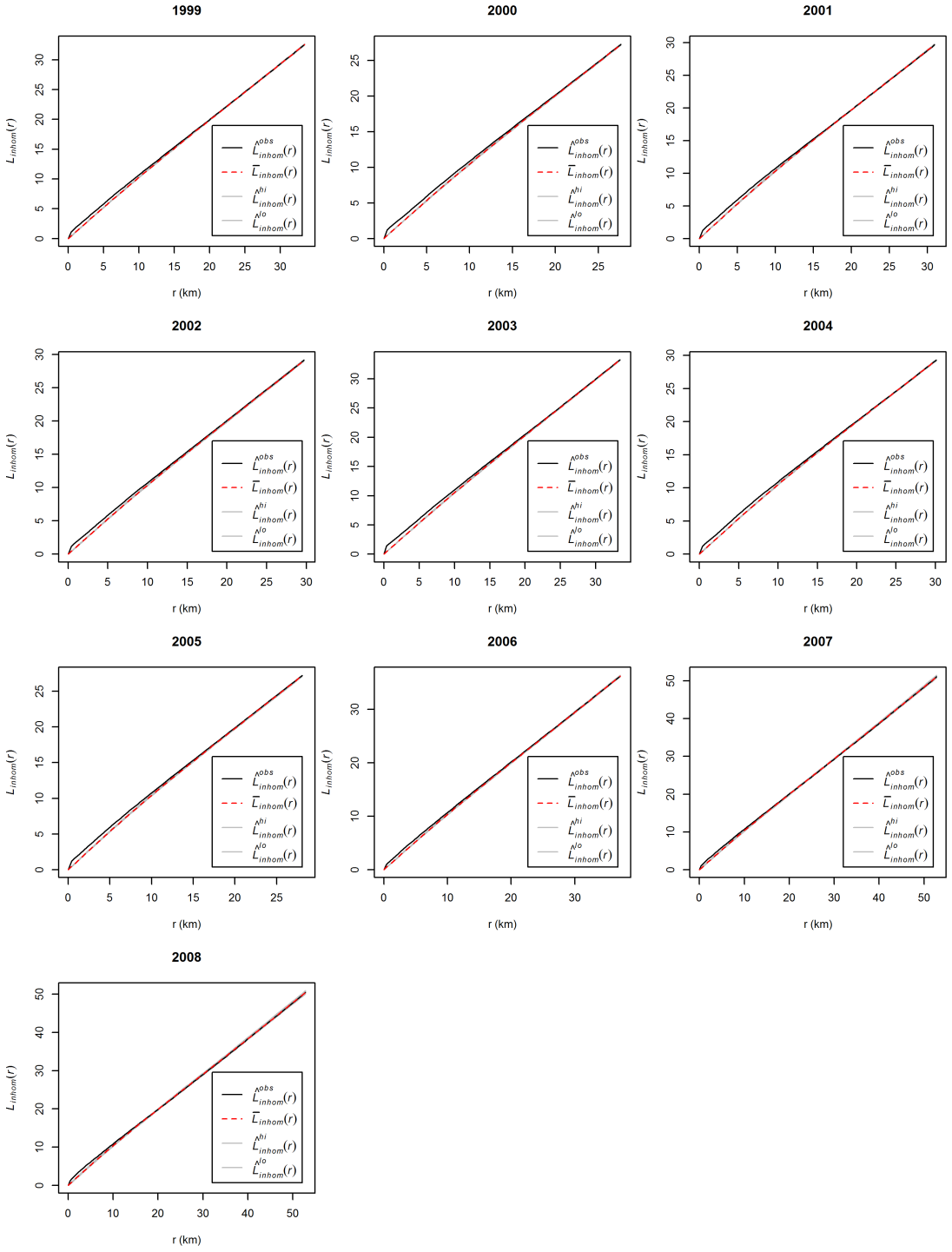


FIGURE B.21: Inhomogeneous L-test for wildfire spatial patterns by year

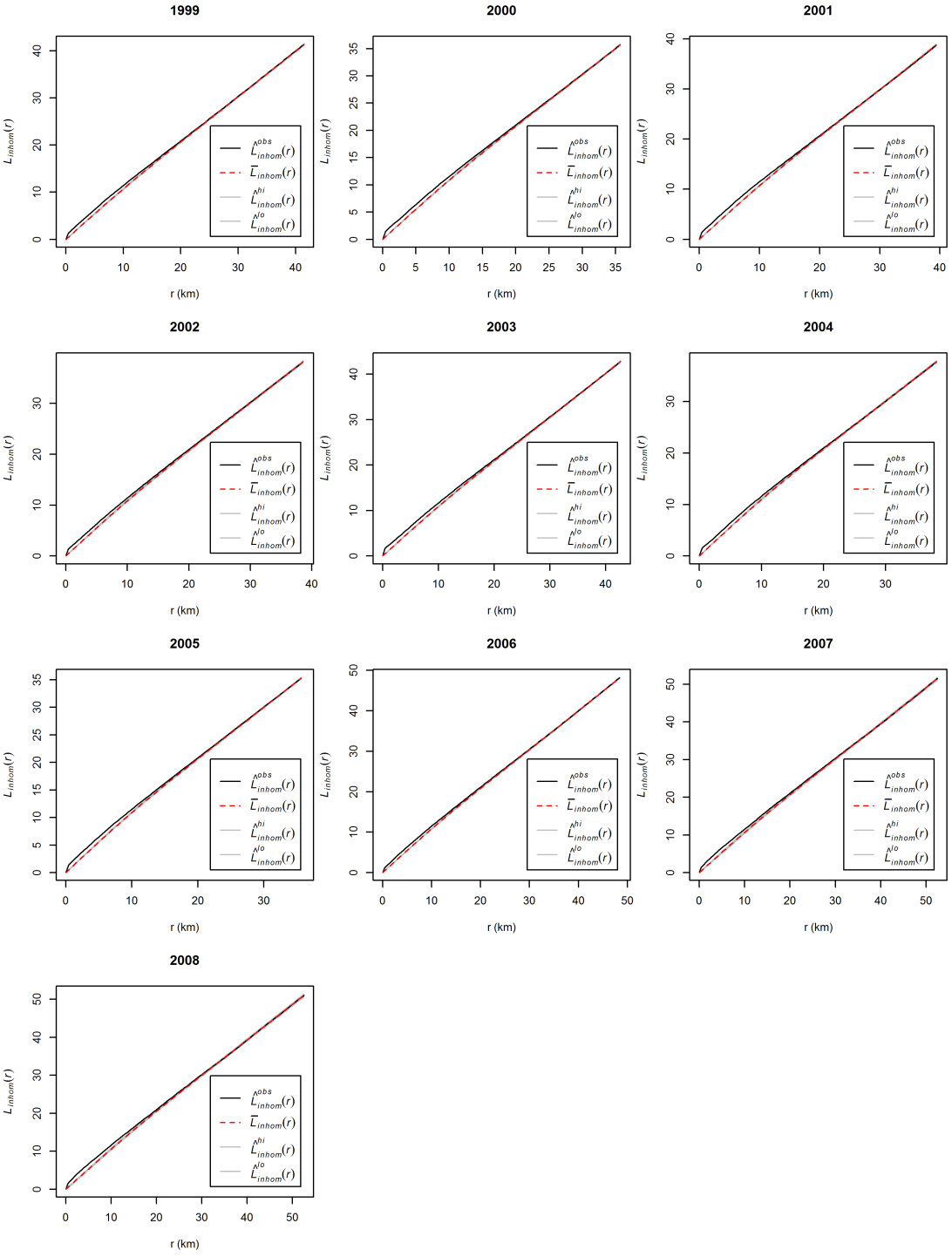


FIGURE B.22: Inhomogeneous L-test for the spatial patterns of small ($S < 1$ ha) wildfires by year

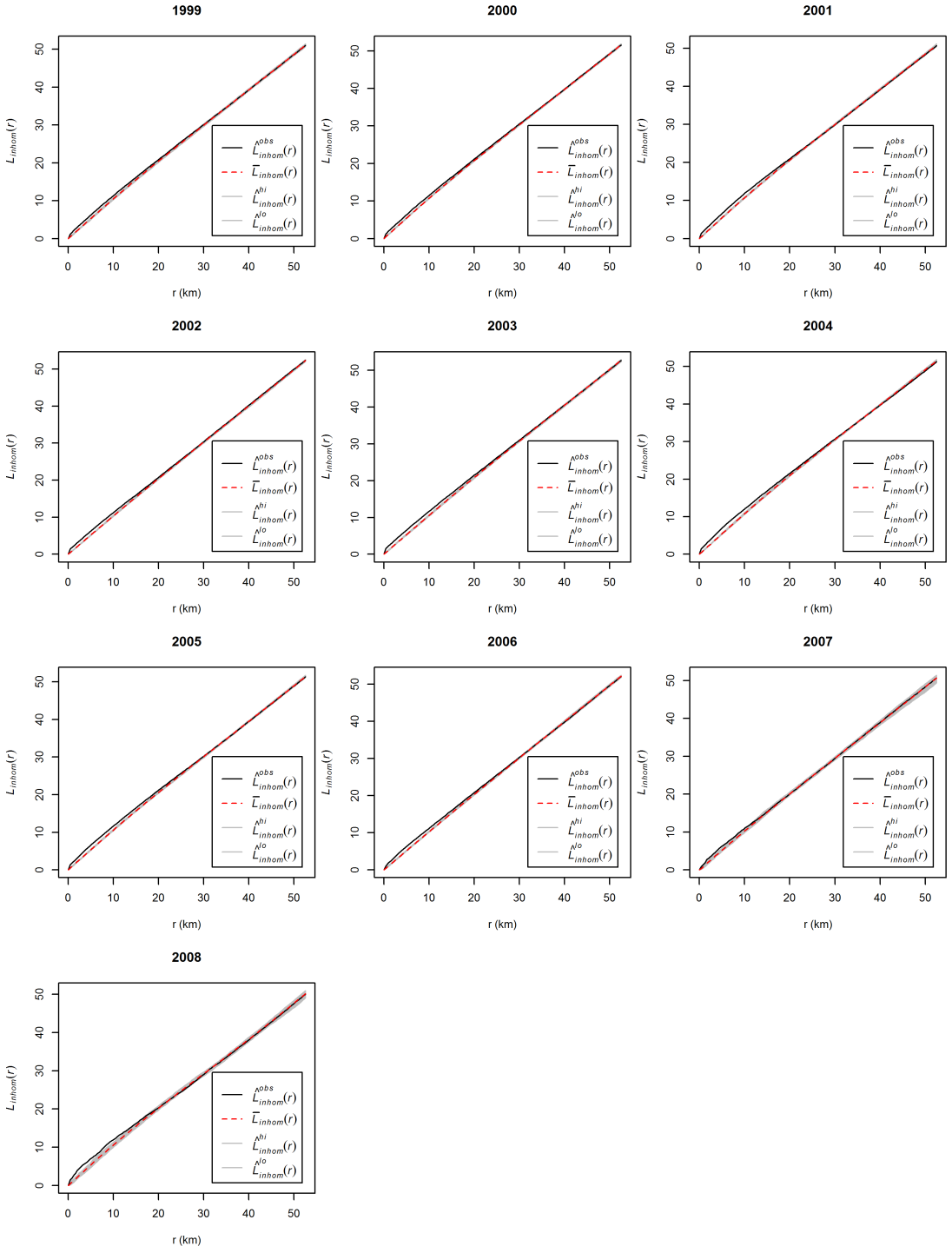


FIGURE B.23: Inhomogeneous L-test for the spatial patterns of regular ($1 \leq S < 25$ ha) wildfires by year

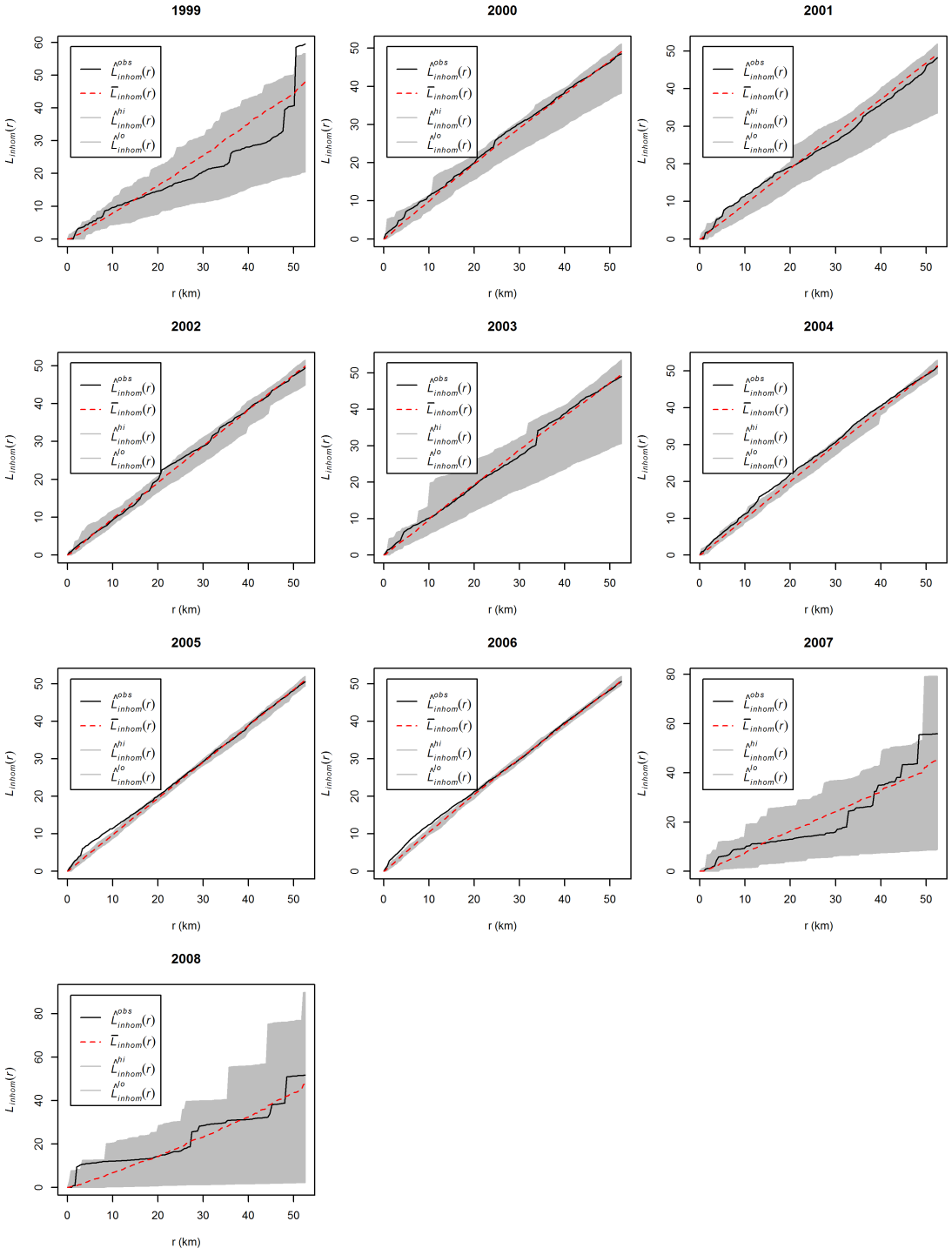


FIGURE B.24: Inhomogeneous L-test for the spatial patterns of large ($S \geq 25$ ha) wildfires by year

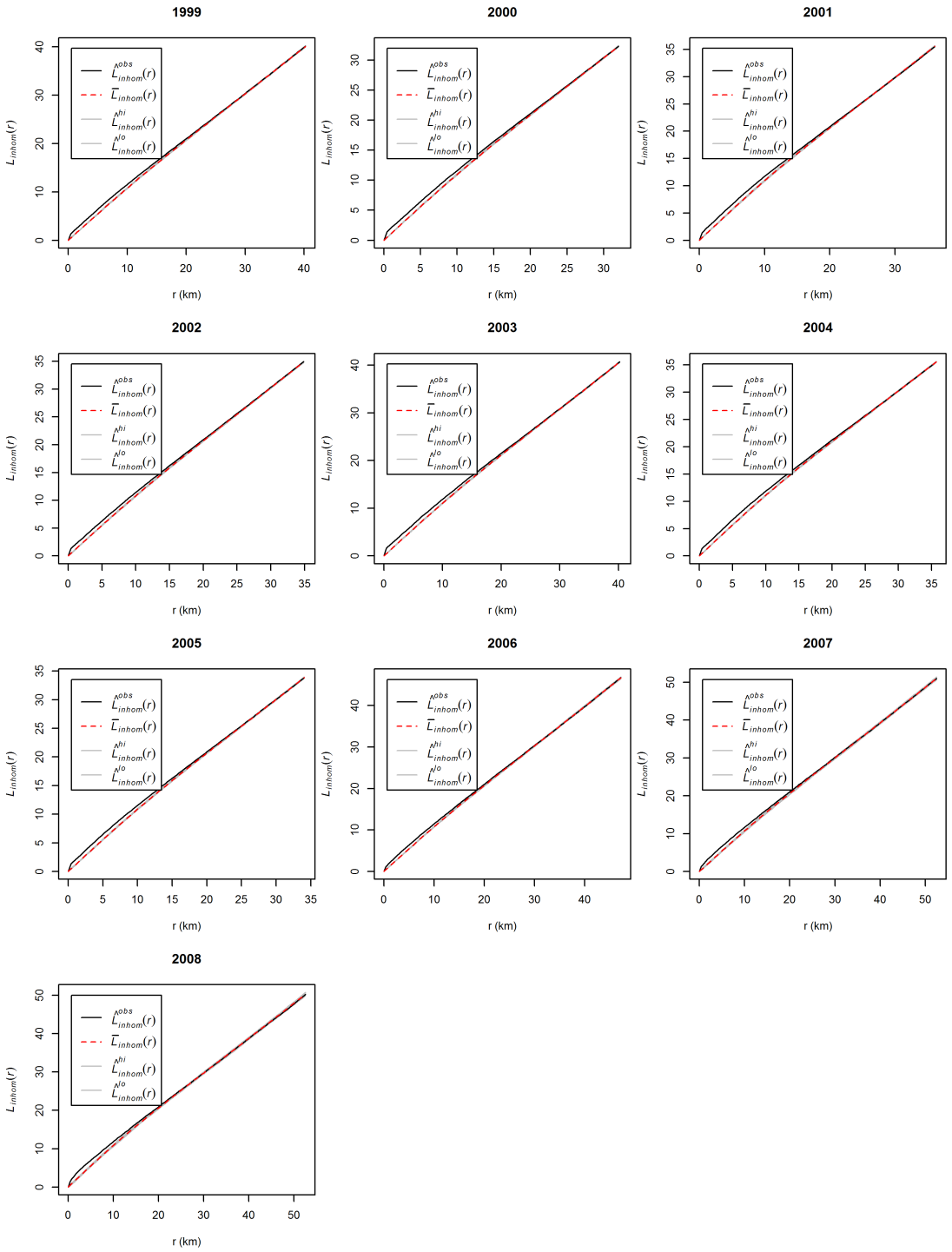


FIGURE B.25: Inhomogeneous L-test for the spatial patterns of arson wildfires by year

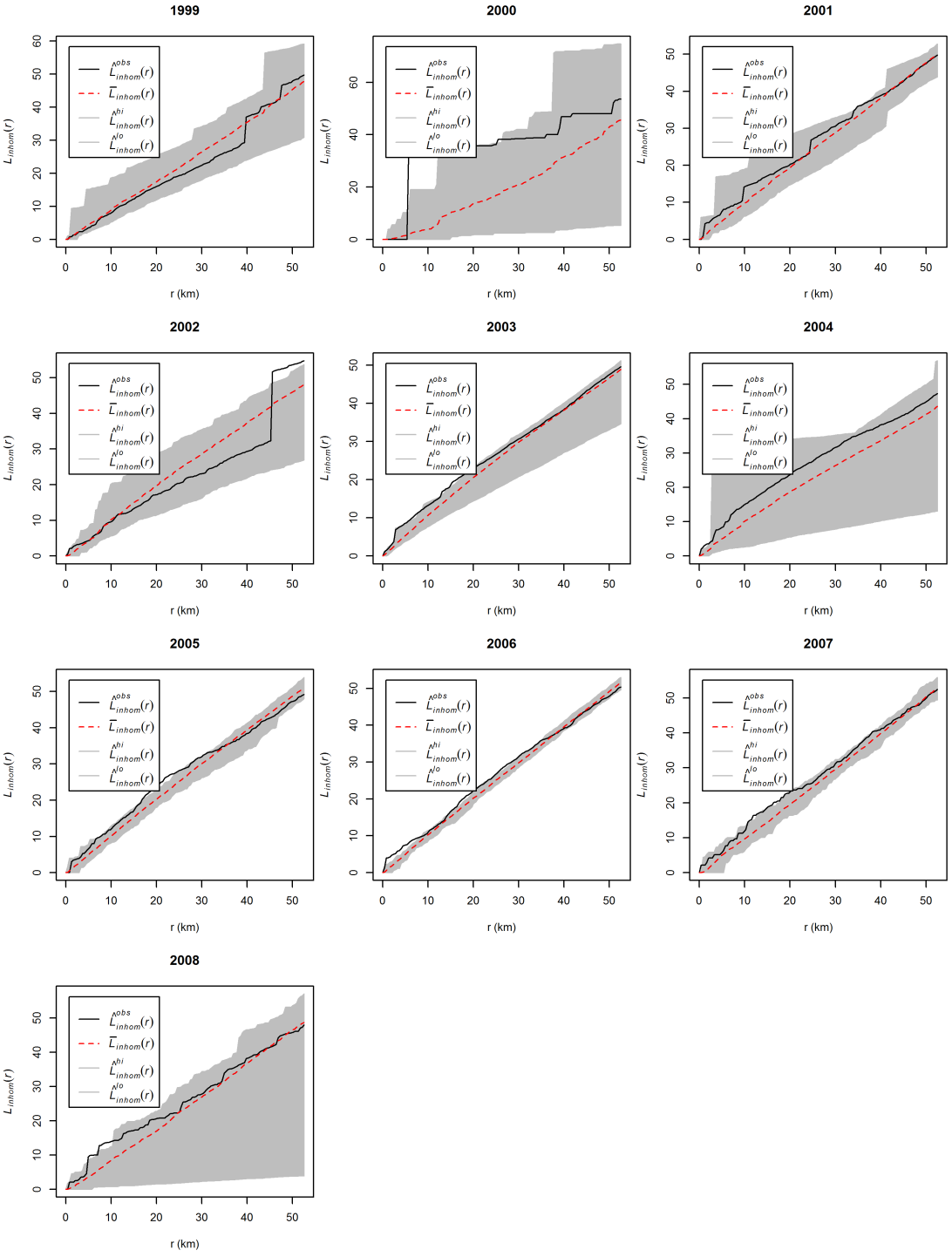


FIGURE B.26: Inhomogeneous L-test for the spatial patterns of natural wildfires by year

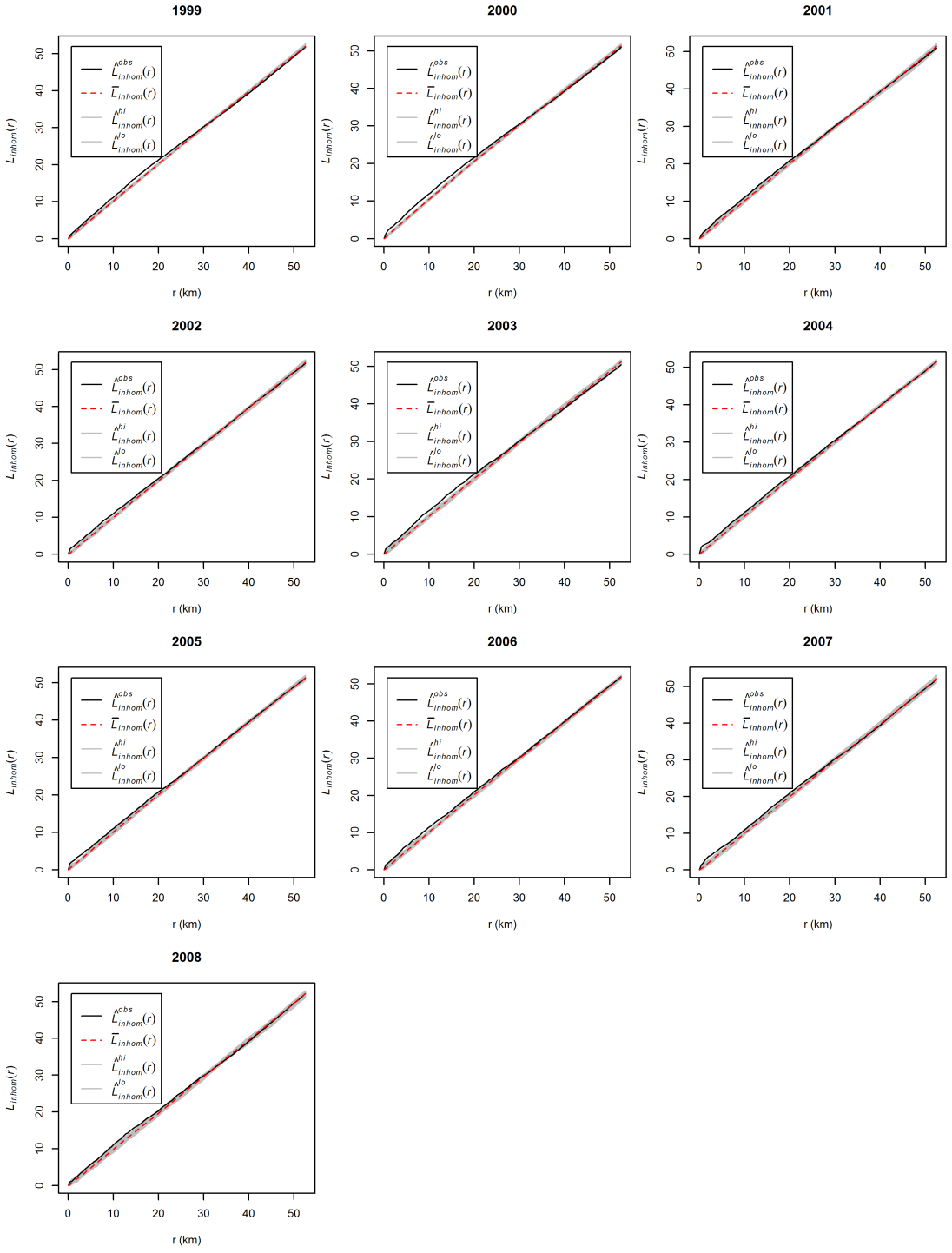


FIGURE B.27: Inhomogeneous L-test for the spatial patterns of wildfires caused by negligence by year

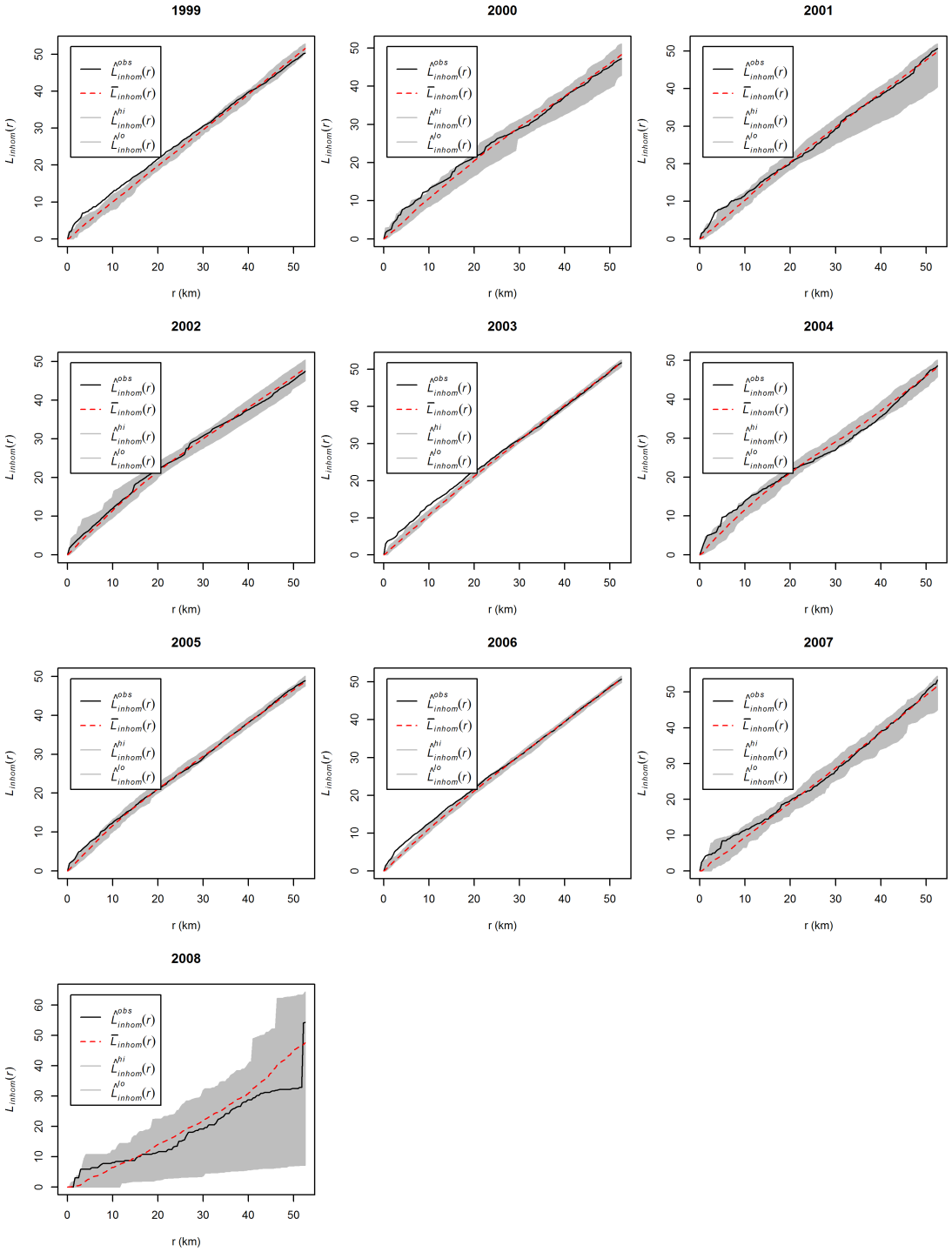


FIGURE B.28: Inhomogeneous L-test for the spatial patterns of reproductions by year

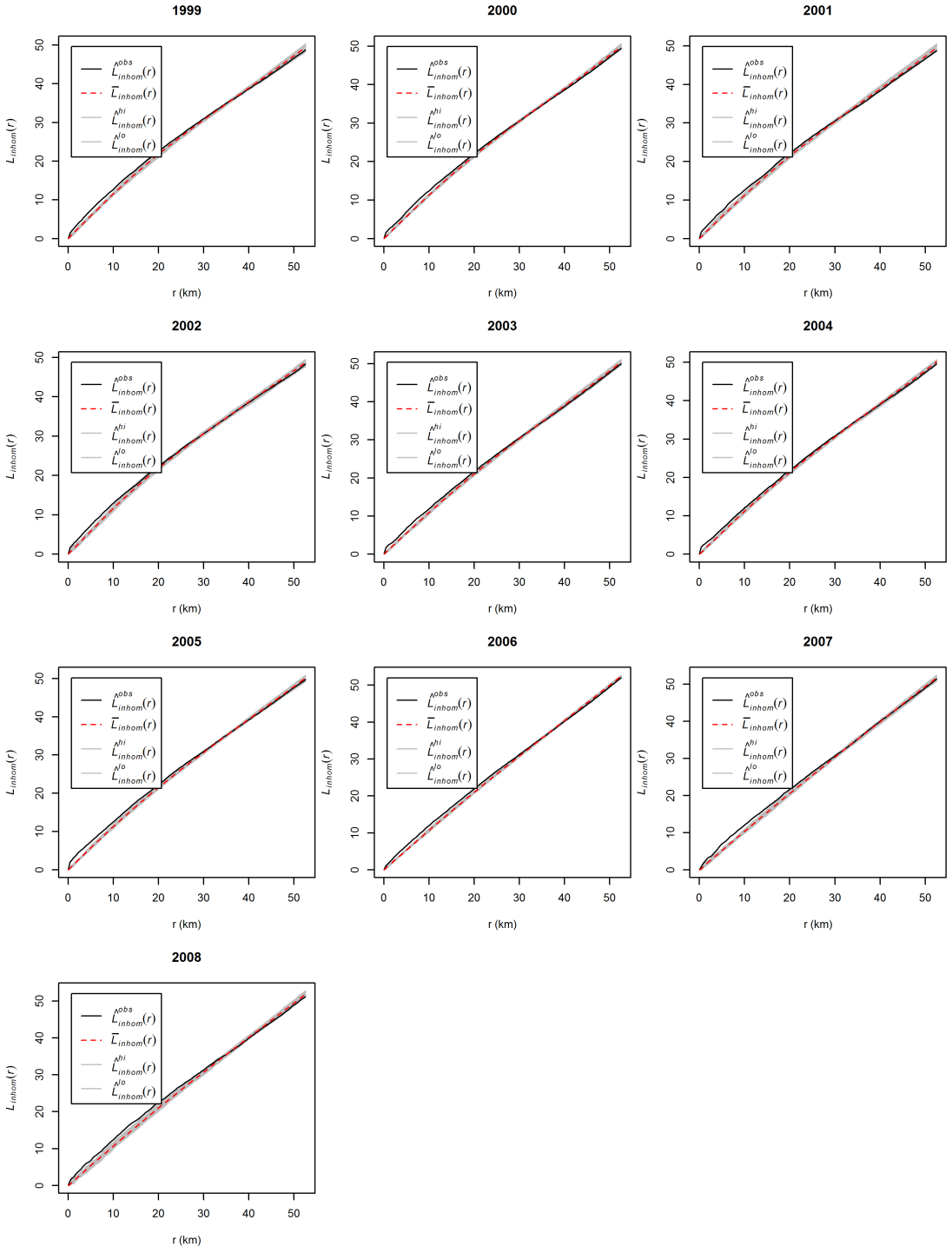


FIGURE B.29: Inhomogeneous L-test for the spatial patterns of wildfires with unknown cause by year

Appendix C

Supporting information for Chapter 6

C.1 Asymptotic properties of the kernel log-ratio estimator

The kernel estimator of the log-ratio function is

$$\hat{\rho}(x,t) = \log\left(\frac{\hat{f}_{h_s,h_t}(x,t)}{\hat{g}_{h_s}(x)}\right)$$

where $hat f_{h_s,h_t}(x,t)$ and $\hat{g}_{h_s}(x)$ are the kernel estimators of the spatio-temporal and spatial densities of event locations, respectively. In order to estimate the bias and variance of the log-ratio estimator we define the relative error terms $\epsilon_f(x,t)$ and $\epsilon_g(x)$:

$$\epsilon_f(x,t) = \frac{\hat{f}_{h_s,h_t}(x,t) - f(x,t)}{f(x,t)} \Rightarrow \hat{f}_{h_s,h_t}(x,t) = f(x,t)(1+\epsilon_f(x,t))$$

$$\epsilon_g(x) = \frac{\hat{g}_{h_s}(x) - g(x)}{g(x)} \Rightarrow \hat{g}_{h_s}(x) = g(x)(1+\epsilon_g(x))$$

Therefore, assuming that the error terms are small, the log-ratio estimator becomes:

$$\begin{aligned}\hat{\rho}(x,t) = \log\left[\frac{f(x,t)(1+\epsilon_f(x,t))}{g(x)(1+\epsilon_g(x))}\right] &= \rho(x,t) + \log\left(1+\epsilon_f(x,t)\right) - log\left(1+\epsilon_g(x)\right) \\ &= \hat{\rho}(x,t) + \epsilon_f(x,t) - \epsilon_g(x) + o(\epsilon_f^2 + \epsilon_g^2)\end{aligned}$$

Overlooking the high order error terms, we obtain the following expression for the mean of $\hat{\rho}(x,t)$:

$$E\left[\hat{\rho}(x,t)\right] = \rho(x,t) + E\left[\epsilon_f(x,t)\right] - E\left[\epsilon_g(x)\right]$$

where

$$E\left[\epsilon_f(x,t)\right] = \frac{E\left[\hat{f}_{h_s,h_t}(x,t) - f(x,t)\right]}{f(x,t)} = \frac{Bias\left(\hat{f}_{h_s,h_t}(x,t)\right)}{f(x,t)}$$
$$= \frac{1}{2}\left[\frac{\mu_2(k_s)h_s^2\nabla_x^2 f(x,t) + \mu_2(k_t)h_t^2\nabla_t^2 f(x,t)}{f(x)}\right] + o\left(h_{s,1}^2 + h_{s,2}^2 + h_t^2\right)$$

and

$$E\left[\epsilon_g(x)\right] = \frac{1}{2}\mu_2(k_s)h_s^2\nabla^2 g(x) + o(h_s^2)$$

This yields the following expression for the expectation of the log-ratio estimator

$$E\left[\hat{\rho}(x,t)\right] = \rho(x,t) + \frac{1}{2}\left[\frac{\mu_2(k_s)h_s^2\nabla_x^2 f(x,t) + h_t^2\mu_2(k_t)\nabla_t^2 f(x,t)}{f(x,t)} - \frac{\mu_2(k_s)h_s^2\nabla^2 g(x)}{g(x)}\right]$$
$$+ o\left(h_{s,1}^2 + h_{s,2}^2 + h_t^2\right)$$

leading to expression (6.2.6).

The variance of the log-ratio estimator is:

$$Var\left(\hat{\rho}(x,t)\right) = Var(\epsilon_f) + Var(\epsilon_g) + Cov(\epsilon_f, \epsilon_g)$$

given that the error terms are assumed to be independent, we have:

$$Var\left(\hat{\rho}(x,t)\right) = \frac{Var(\hat{f}_{h_s,h_t}(x,t))}{f(x,t)^2} + \frac{Var(\hat{g}_{h_s}(x,))}{g(x)^2}$$
$$= \frac{R(k_s)R(k_t)}{nh_{s,1}h_{s,2}h_t f(x,t)} + \frac{R(k_s)}{nh_{s,1}h_{s,2}g(x)} + o\left((nh_{s,1}h_{s,2}\left(1+h_t\right))^{-1}\right)$$

C.2 Details for the LSCV bandwidth selector

As shown in expression (6.2.10), the least-squares cross-validation bandwidth selector aims to minimize,

$$LSCV\left(h_s,h_t\right)=\int_{W\times T}\left(\gamma(f(x,t),g(x)\right)^2dxdt \\ -2\int_{W\times T}\gamma(\hat{f}_{h_s,h_t}(x,t),\hat{g}_{h_s}(x))\gamma(f(x,t),g(x))dxdt$$

where $\gamma\left(f,g\right)$ is a smooth functional of the spatio-temporal and spatial densities.

Applying a first-order Taylor expansion to $\gamma(f,g)$

$$\gamma(f,g)=\gamma\left(\hat{f},\hat{g}\right)+\left(f-\hat{f}\right)\frac{\partial\gamma}{\partial\hat{f}}+\left(g-\hat{g}\right)\frac{\partial\gamma}{\partial\hat{g}}$$

and substituting in (6.2.10) we obtain

$$LSCV\left(h_s,h_t\right)=\int_{W\times T}\gamma\left(f(x,t),g(x)\right)^2dxdt \\ -2\int_{W\times T}\gamma(\hat{f}(x,t),\hat{g}(x))\left(\gamma(\hat{f}(x,t),\hat{g}(x))\left(f(x,t)-\hat{f}(x,t)\right)\frac{\partial\gamma}{\partial\hat{f}}+(g(x)-\hat{g}(x))\frac{\partial\gamma}{\partial\hat{g}}\right)dxdt \\ =-\int_{W\times T}\gamma\left(\hat{f}(x,t),\hat{g}(x)\right)^2dxdt+2\int_{W\times T}A_1(\hat{f},\hat{g})\hat{f}(x,t)dxdt+2\int_{W\times T}A_2(\hat{f},\hat{g})\hat{g}(x)dxdt \\ -2\int_{W\times T}A_1(\hat{f},\hat{g})f(x,t)dxdt-2\int_{W\times T}A_2(\hat{f},\hat{g})g(x)dxdt$$

where $\hat{f}=\hat{f}_{h_s,h_t}$ and $\hat{g}=\hat{g}_{h_s}$, and $A_i(c_1,c_2)=\gamma(c_1,c_2)\frac{\partial\gamma}{\partial c_i}$, for $i=1,2$. The last two terns in the previous expression are means with respect to the unknown densities f and g, which can be estimated by cross-validation, yielding

$$\int_{W\times T}A_1\left(\hat{f},\hat{g}\right)f(x,t)dxdt=n^{-1}\sum_{i=1}^{n}\gamma\left(\hat{f}^{-i}(x_i,t_i),\hat{g}(x_i)\right)\frac{\partial\gamma}{\partial\hat{f}}$$

$$\int_{W\times T}A_2\left(\hat{f},\hat{g}\right)f(x,t)dxdt=n^{-1}\sum_{i=1}^{n}\gamma\left(\hat{f}(x_i,t_i),\hat{g}^{-i}(x_i)\right)\frac{\partial\gamma}{\partial\hat{g}}$$

where $\hat{f}^{-i}$ and $\hat{g}^{-i}$ denote the leave-one out estimators of f and g, defined as follows

$$\hat{f}^{-i}(x,t)=\frac{\sum\limits_{j\neq i}k_s\left(H_s^{-1/2}\left(x-x_j\right)\right)k_t\left(h_t^{-1}\left(t-t_j\right)\right)}{(n-1)p_{h_s,h_t}(x,t)|H_s|^{1/2}h_t}$$

$$\hat{g}^{-i}(x) = \frac{\sum_{j \neq i} k_s\left(H^{-1/2}\left(x - X_j\right)\right)}{(n-1)p_{H_s}(x)|H_s|^{1/2}}$$

where $H_s = h_s^2 I_2$.

Replacing $\gamma(f, g)$ by $\rho(x, t) = \log(f/g)$ and taking into account that $\dfrac{\partial \rho(x,t)}{\partial \hat{f}} = \dfrac{1}{\hat{f}}$ and $\dfrac{\partial \rho(x,t)}{\partial \hat{g}} = -\dfrac{1}{\hat{g}}$, the second and third terms in the previous expression cancel and we obtain expression (3.4.3).

$$LSCV(h_s, h_t) = -\int_T \int_W \hat{\rho}_{h_s,h_t}(x,t)^2 dxdt - 2n^{-1}\sum_{i=1}^{n} \log\left(\frac{\hat{f}^{-i}_{h_s,h_t}(x_i,t_i)}{\hat{g}_{h_s}(x_i)}\right)\left(\hat{f}^{-i}_{h_s,h_t}(x_i,t_i)\right)^{-1}$$
$$+ 2n^{-1}\sum_{j=1}^{n} \log\left(\frac{\hat{f}_{h_s,h_t}(x_j,t_j)}{\hat{g}^{-j}_{h_s}(x_j)}\right)\left(\hat{g}^{-j}_{h_s}(x_j)\right)^{-1}$$

C.3 Details for the simulation studies conducted in Section 6.4

FIGURE C.1: Spatial pattern (left), spatial intensity (center) and temporal density (right) of simulated patterns with first-order intensity defined in (6.4.1) for $\epsilon = 0, 0.1, 0.3, 0.5$

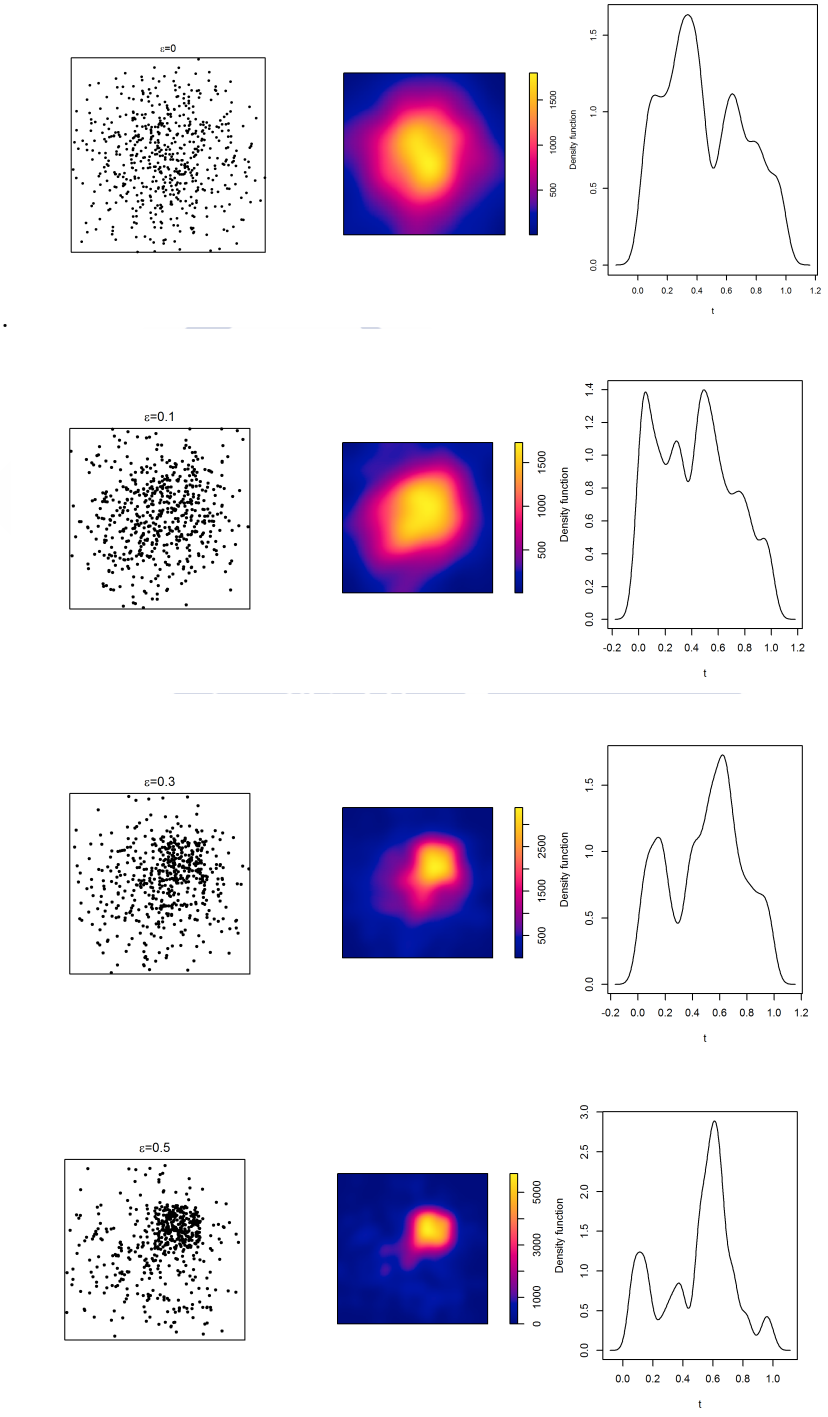


FIGURE C.2: Kernel estimator of the spatio-temporal intensity function applying binning with a resolution of 20 bins in each dimension and plug-in bandwidth. Separable point patterns with first-order intensity defined in (6.4.1).

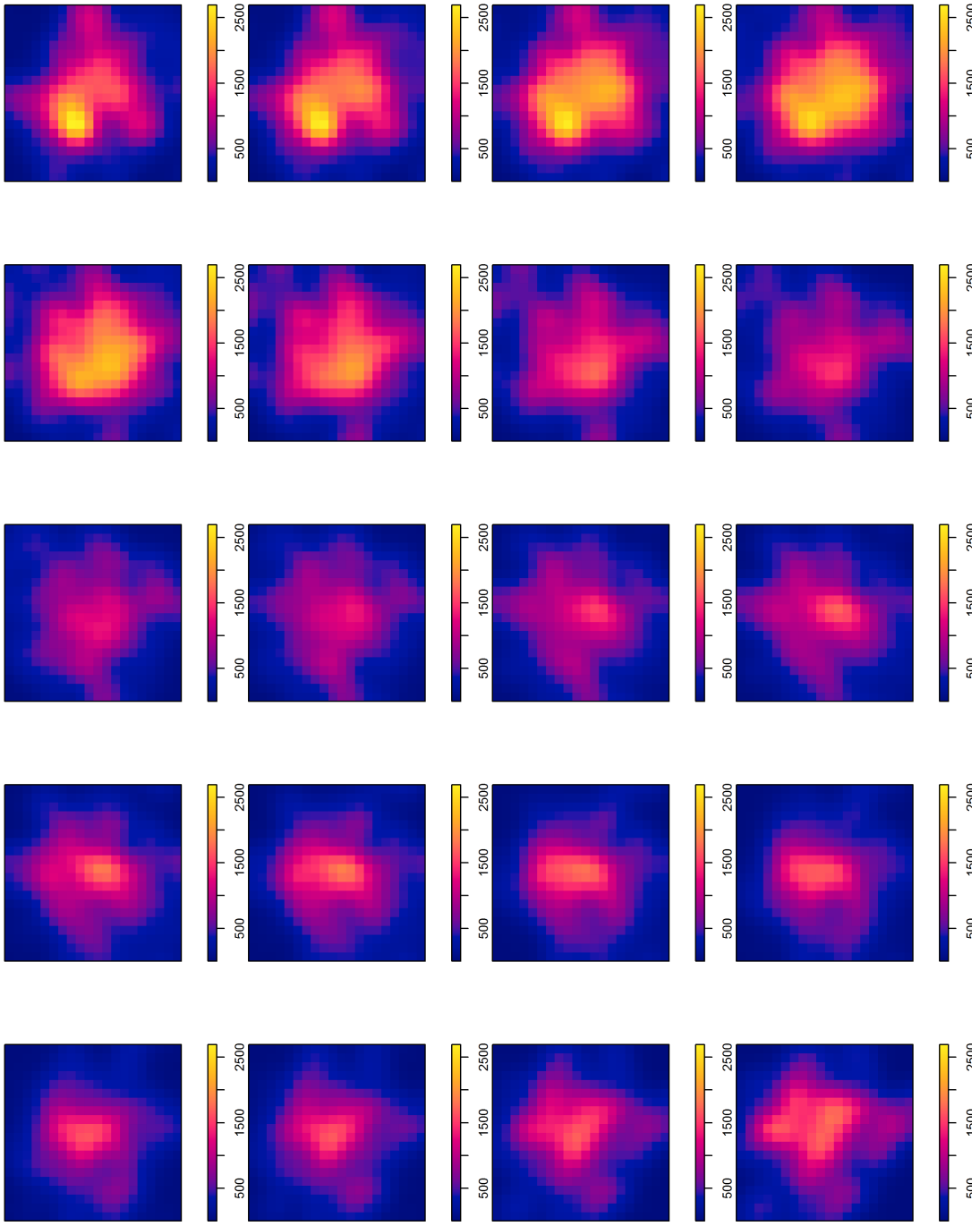


FIGURE C.3: Kernel estimator of the spatio-temporal intensity function applying binning with a resolution of 20 bins in each dimension and plug-in bandwidth. Nonseparable point patterns with first-order intensity defined in (6.4.1) and $\epsilon = 0.2$.

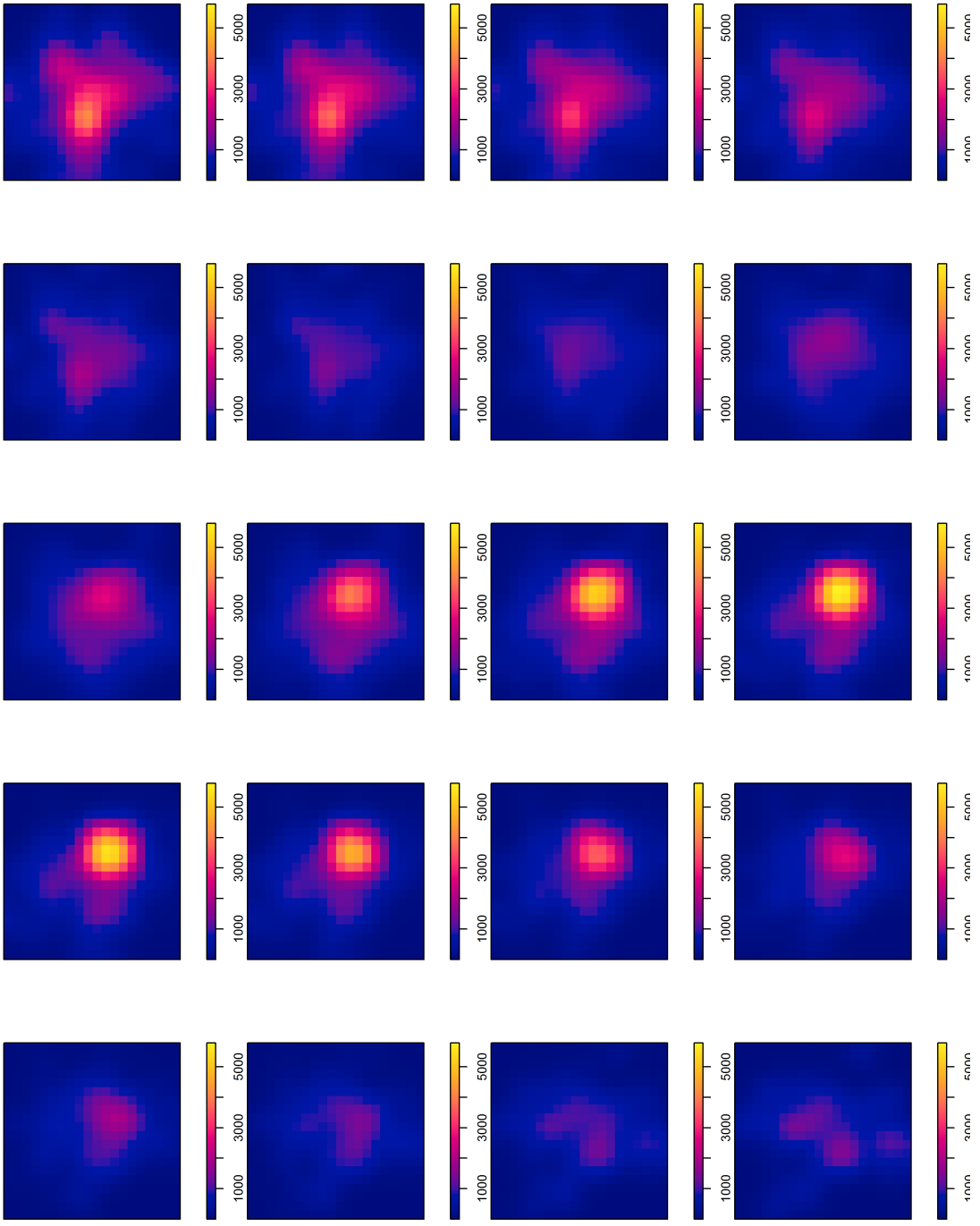


FIGURE C.4: Spatial pattern (left), spatial intensity (center) and temporal density (right) of simulated patterns with first-order intensity defined in expression (6.4.2) for $\epsilon = 0, 0.1, 0.3, 0.5$

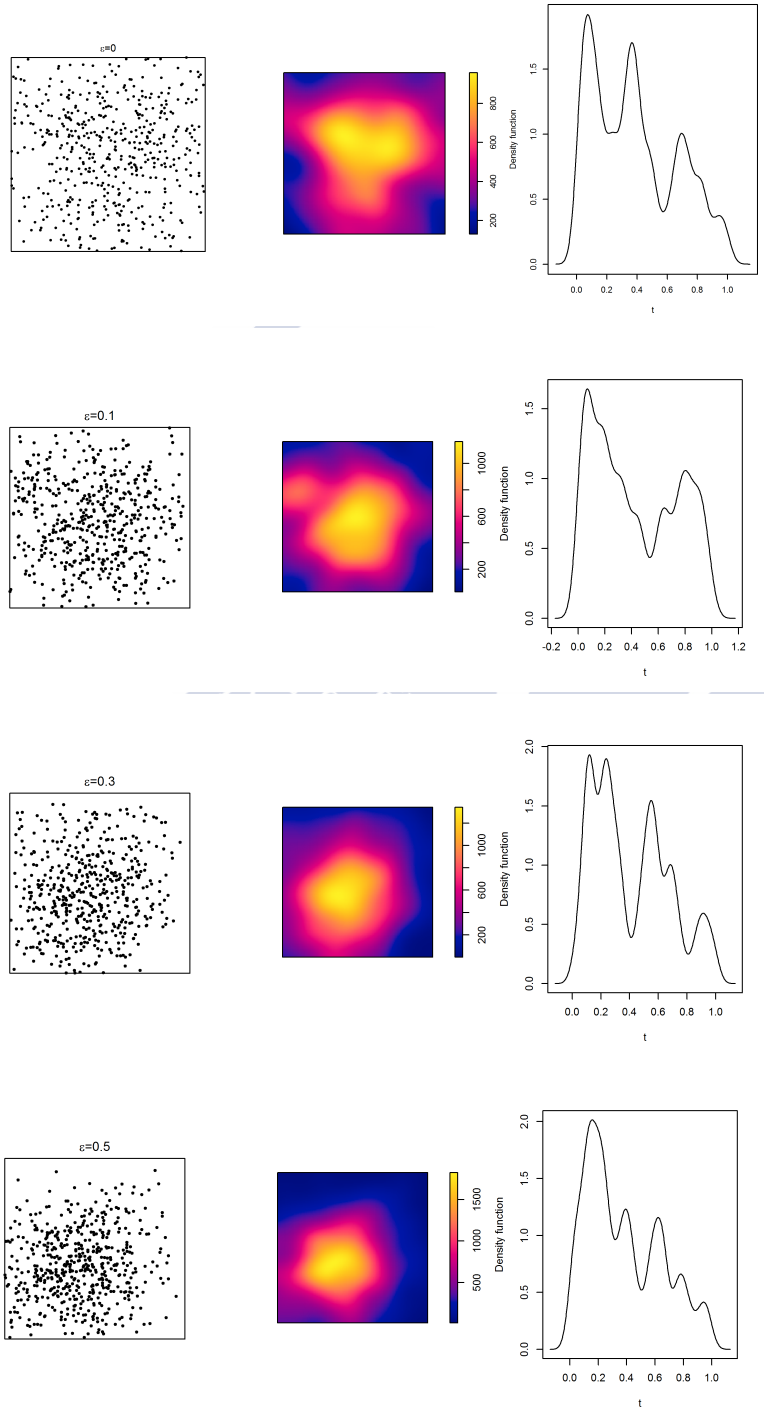


FIGURE C.5: Kernel estimator of the spatio-temporal intensity function applying binning with a resolution of 20 bins in each dimension and plug-in bandwidth. Separable point patterns with first-order intensity defined in expression (6.4.2).

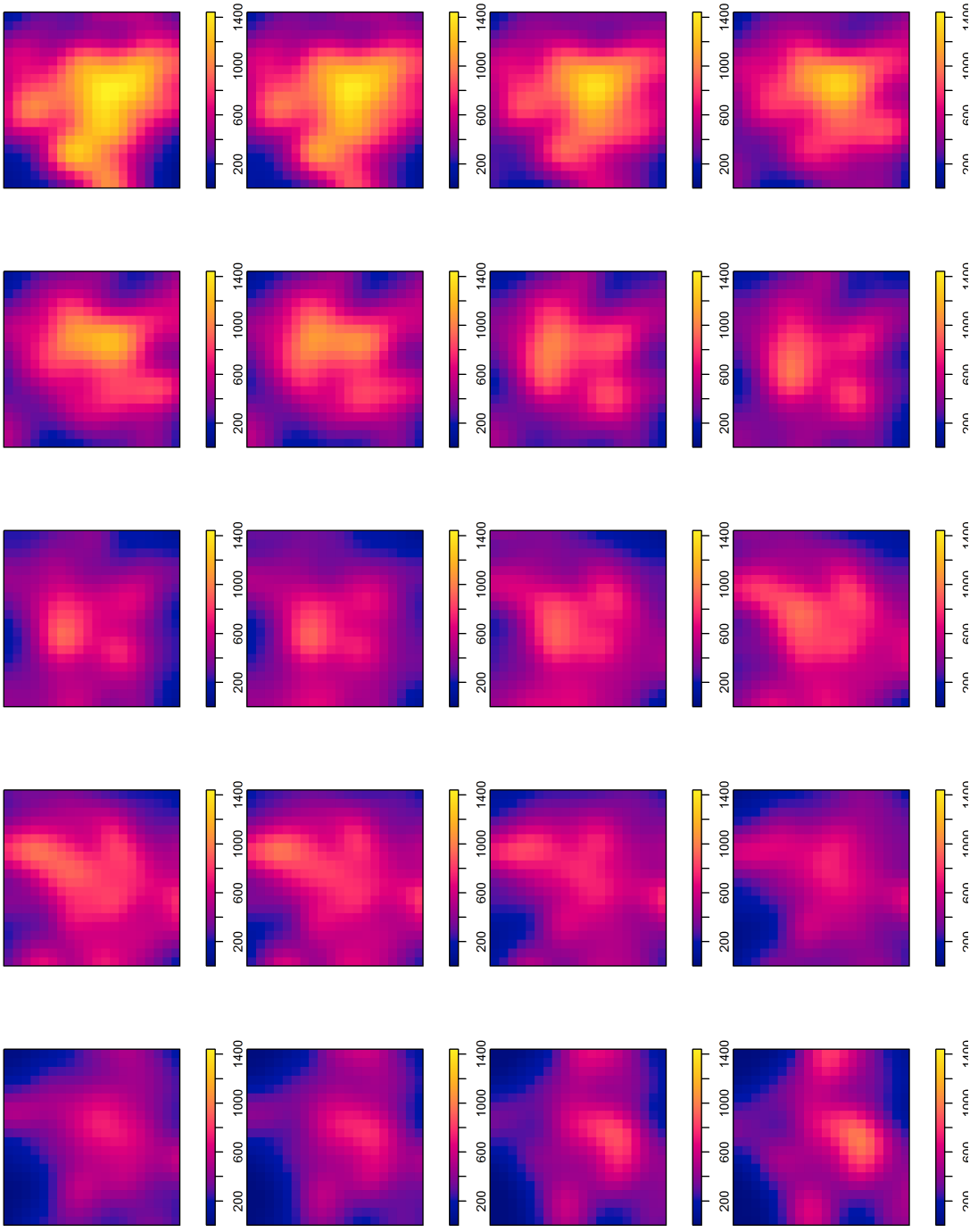
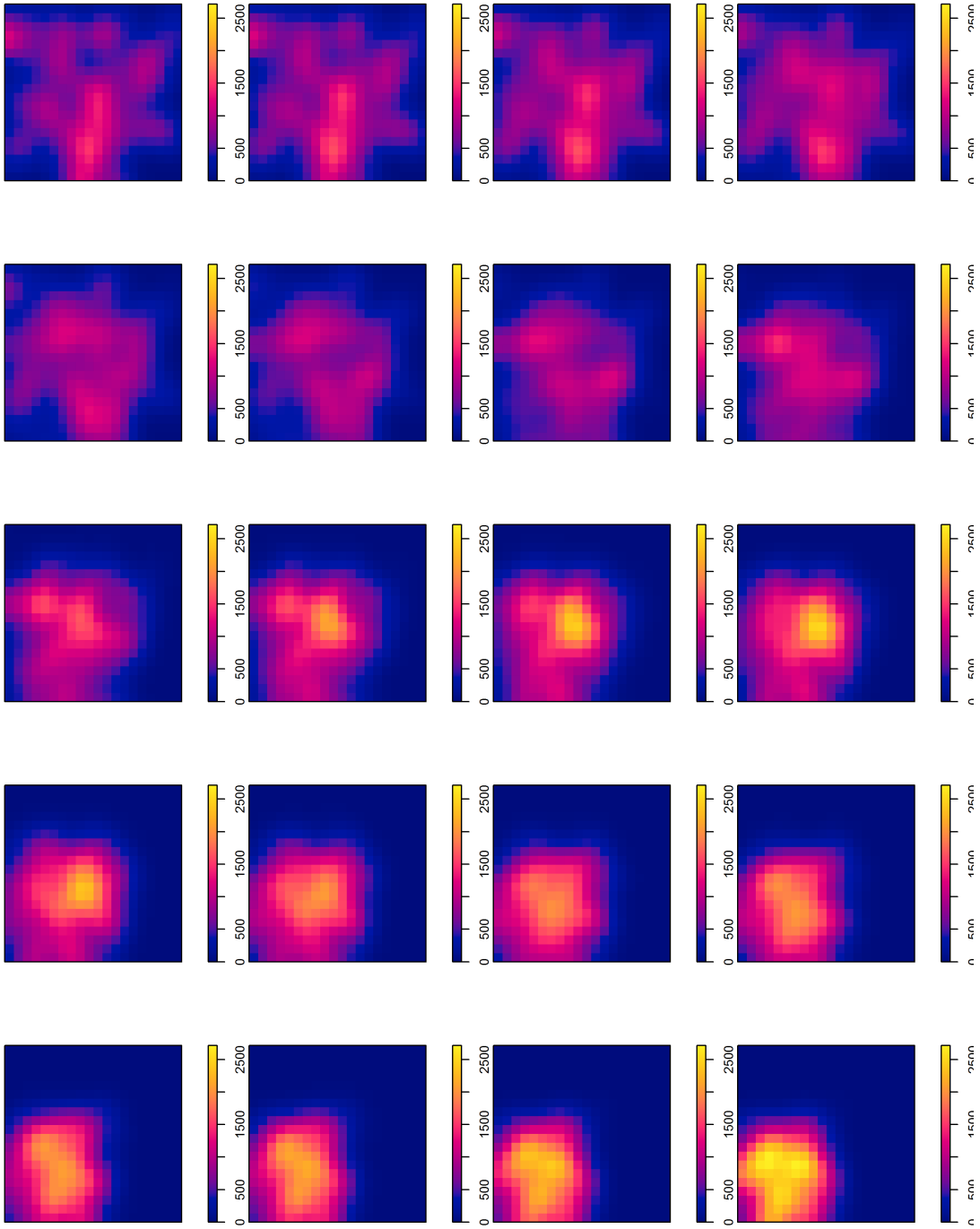


FIGURE C.6: Kernel estimator of the spatio-temporal intensity function applying binning with a resolution of 20 bins in each dimension and plug-in bandwidth. Nonseparable point patterns with first-order intensity defined in expression (6.4.2) and $\epsilon = 0.5$.



Comparison between calibration procedures

FIGURE C.7: Power comparison for simulated patterns with intensity given in expression (6.4.1). Null distribution of the test statistic estimated by the permutation test (top) and χ^2 approximation (bottom).

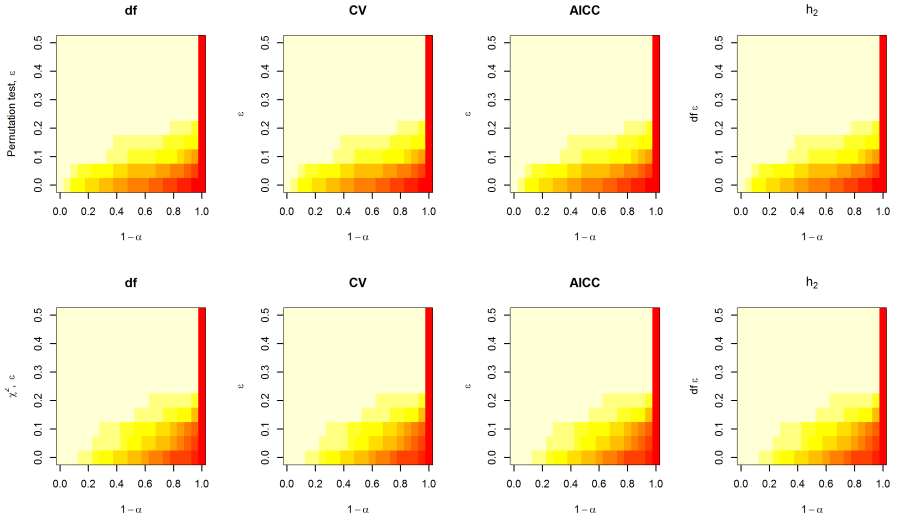
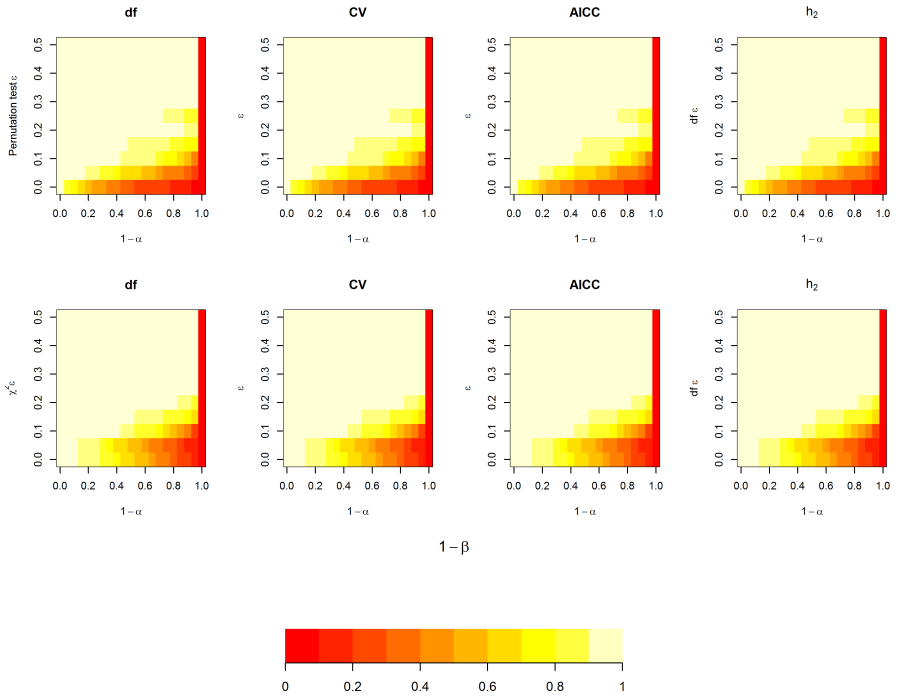


FIGURE C.8: Power comparison for simulated patterns with intensity given in expression (6.4.2). See details in the caption of Figure C.7

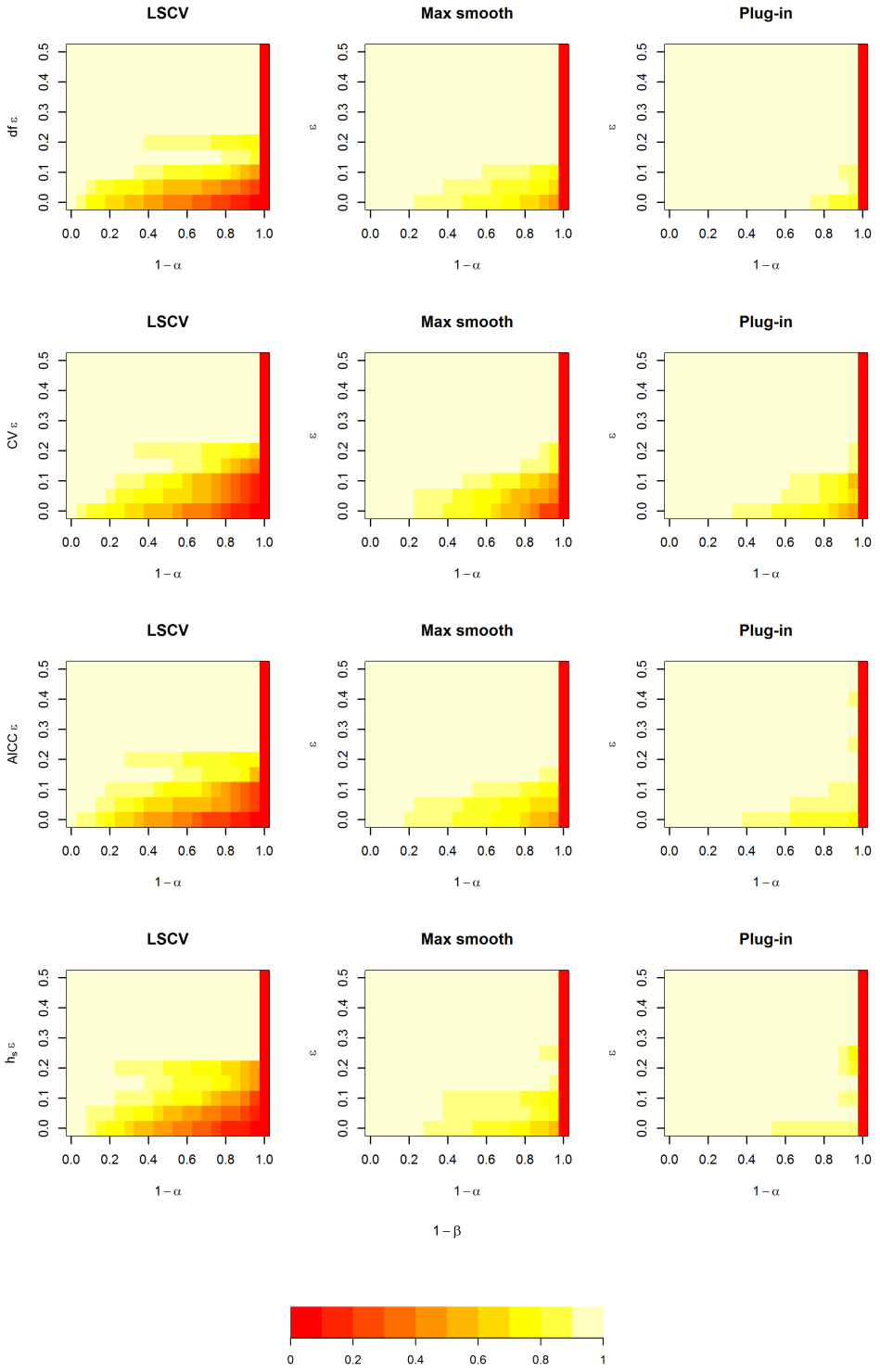


Effect of the bandwidth parameters on the test

		$\alpha = 0.05$				$\alpha = 0.01$			
ϵ	Selector	df	CV	AICC	H_S	df	CV	AICC	H_S
0.00	LSCV	0.090	0.075	0.070	0.055	0.030	0.030	0.025	0.010
	Max smooth	0.675	0.500	0.555	0.755	0.540	0.350	0.410	0.635
	Plug-in	0.870	0.720	0.745	0.945	0.740	0.600	0.680	0.890
0.05	LSCV	0.220	0.195	0.120	0.090	0.090	0.085	0.060	0.030
	Max smooth	0.750	0.570	0.615	0.860	0.630	0.470	0.525	0.770
	Plug-in	0.880	0.765	0.810	0.965	0.795	0.640	0.735	0.925
0.10	LSCV	0.320	0.350	0.265	0.145	0.210	0.220	0.170	0.055
	Max smooth	0.830	0.750	0.745	0.895	0.765	0.635	0.645	0.825
	Plug-in	0.930	0.835	0.880	0.980	0.890	0.790	0.850	0.960
0.15	LSCV	0.555	0.535	0.460	0.360	0.385	0.400	0.330	0.230
	Max smooth	0.935	0.870	0.860	0.975	0.890	0.825	0.795	0.925
	Plug-in	0.985	0.950	0.960	0.995	0.965	0.910	0.935	0.990
0.20	LSCV	0.795	0.760	0.730	0.535	0.685	0.675	0.630	0.395
	Max smooth	0.980	0.960	0.960	0.995	0.950	0.930	0.925	0.975
	Plug-in	0.990	0.985	0.990	0.985	0.980	0.965	0.985	0.980
0.25	LSCV	0.965	0.965	0.960	0.930	0.955	0.960	0.955	0.885
	Max smooth	1.000	0.995	1.000	0.995	0.990	0.990	0.975	0.990
	Plug-in	0.995	0.990	0.990	0.990	0.985	0.985	0.990	0.985
0.30	LSCV	0.985	0.965	0.965	0.945	0.975	0.965	0.965	0.930
	Max smooth	1.000	1.000	0.995	1.000	0.995	0.995	0.995	0.995
	Plug-in	1.000	0.995	0.995	0.995	0.995	0.995	0.990	0.995
0.35	LSCV	1.000	1.000	1.000	0.990	1.000	1.000	1.000	0.985
	Max smooth	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
	Plug-in	1.000	1.000	1.000	0.995	0.995	1.000	0.995	0.995
0.40	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Max smooth	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Plug-in	0.995	1.000	1.000	1.000	0.995	0.995	0.995	1.000
0.45	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Max smooth	1.000	1.000	1.000	0.995	1.000	1.000	1.000	0.995
	Plug-in	1.000	1.000	1.000	0.995	1.000	1.000	0.995	0.995
0.50	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Max smooth	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Plug-in	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000

TABLE C.1: Probability of rejection of separability for the spatio-temporal point processes with first-order intensity (6.4.1). Null distribution of F estimated by χ^2-approximation. Bandwidth selector for the kernel log-ratio function (6.2.3) in rows, and bandwidth selector for the kernel regression function (6.3.1) in columns.

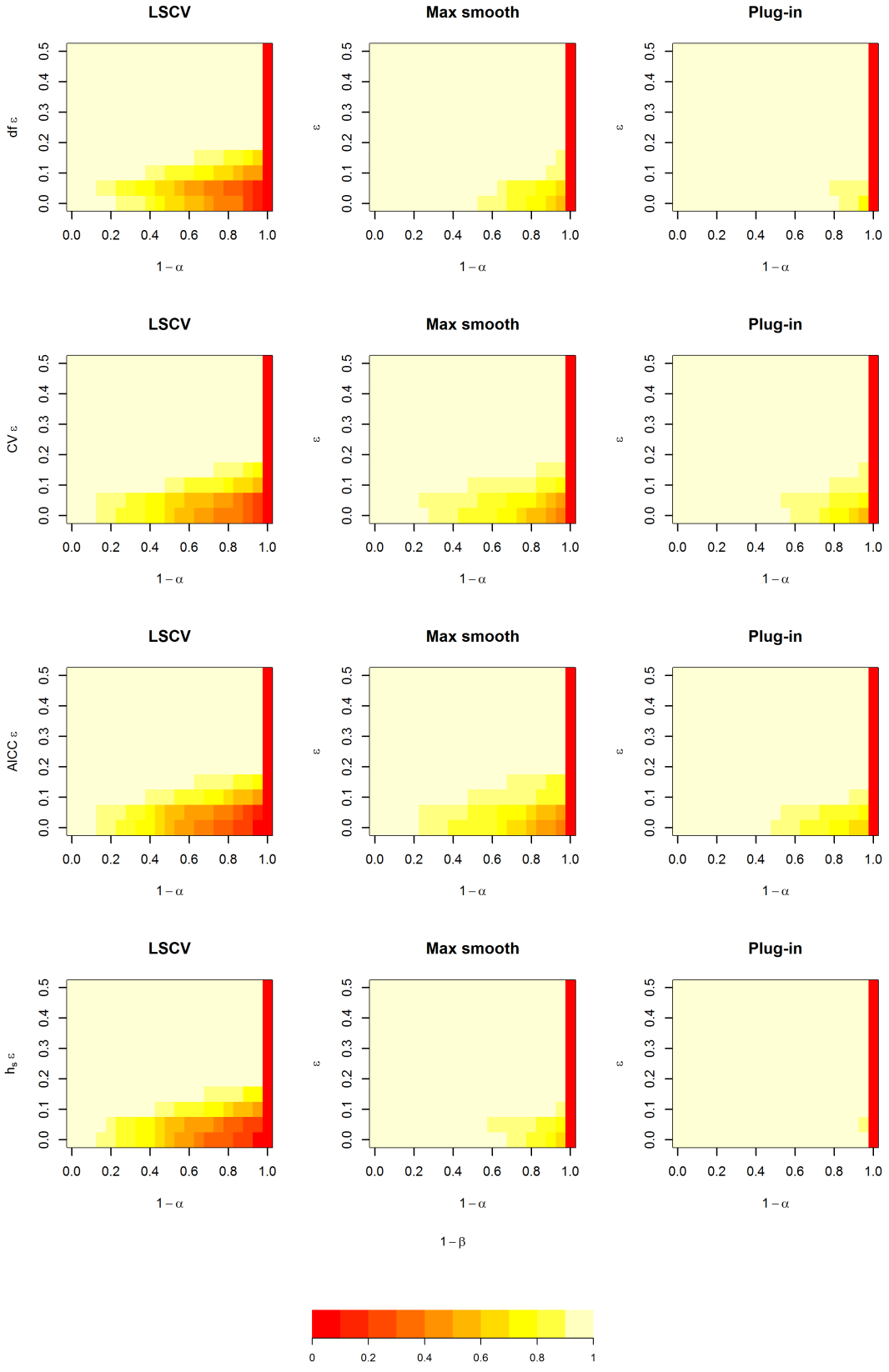
FIGURE C.9: Power comparison for simulated patterns with intensity given in expression (6.4.1). Bandwidth selector for the kernel log-ratio in columns, and bandwidth selector for the kernel regression in rows.



		$\alpha = 0.05$				$\alpha = 0.01$			
ϵ	Selector	df	CV	AICC	H_S	df	CV	AICC	H_S
0.00	LSCV	0.110	0.130	0.100	0.075	0.055	0.075	0.050	0.020
	Max smooth	0.515	0.515	0.415	0.560	0.370	0.350	0.305	0.425
	Plug-in	0.725	0.675	0.630	0.825	0.605	0.560	0.510	0.720
0.50	LSCV	0.100	0.135	0.090	0.065	0.020	0.050	0.025	0.025
	Max smooth	0.440	0.425	0.350	0.470	0.255	0.285	0.240	0.330
	Plug-in	0.630	0.570	0.505	0.720	0.470	0.355	0.365	0.590
0.10	LSCV	0.235	0.440	0.400	0.390	0.085	0.150	0.130	0.135
	Max smooth	0.735	0.715	0.660	0.720	0.470	0.490	0.465	0.500
	Plug-in	0.675	0.670	0.625	0.700	0.535	0.525	0.470	0.595
0.15	LSCV	0.600	0.695	0.660	0.640	0.415	0.515	0.510	0.495
	Max smooth	0.885	0.885	0.865	0.895	0.795	0.815	0.820	0.800
	Plug-in	0.805	0.785	0.790	0.750	0.685	0.655	0.660	0.680
0.20	LSCV	0.780	0.850	0.875	0.860	0.685	0.760	0.750	0.725
	Max smooth	0.965	0.950	0.980	0.935	0.940	0.905	0.955	0.895
	Plug-in	0.860	0.865	0.895	0.770	0.760	0.770	0.815	0.695
0.25	LSCV	0.930	0.945	0.950	0.930	0.870	0.915	0.915	0.905
	Max smooth	0.985	0.985	0.990	0.945	0.950	0.960	0.980	0.915
	Plug-in	0.895	0.885	0.945	0.735	0.805	0.770	0.885	0.645
0.30	LSCV	0.955	0.975	0.970	0.965	0.930	0.940	0.945	0.945
	Max smooth	0.995	0.985	0.990	0.975	0.985	0.970	0.990	0.920
	Plug-in	0.855	0.880	0.965	0.645	0.765	0.750	0.900	0.595
0.35	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	0.995
	Max smooth	0.995	0.990	1.000	0.960	0.990	0.980	0.995	0.905
	Plug-in	0.855	0.865	0.960	0.645	0.770	0.735	0.895	0.575
0.40	LSCV	0.995	0.995	0.990	0.995	0.990	0.990	0.990	0.990
	Max smooth	1.000	0.990	1.000	0.945	0.980	0.975	1.000	0.895
	Plug-in	0.885	0.885	0.995	0.575	0.775	0.810	0.925	0.500
0.45	LSCV	1.000	1.000	1.000	1.000	1.000	1.000	1.000	1.000
	Max smooth	1.000	0.985	1.000	0.935	0.985	0.980	1.000	0.880
	Plug-in	0.830	0.900	0.995	0.560	0.715	0.795	0.920	0.510
0.00 0.50	LSCV	0.995	1.000	1.000	1.000	0.995	0.995	0.995	0.995
	Max smooth	0.995	0.985	1.000	0.960	0.995	0.980	1.000	0.915
	Plug-in	0.880	0.915	0.995	0.580	0.765	0.835	0.980	0.550

TABLE C.2: Probability of rejection of separability for the spatio-temporal point processes with first-order intensity (6.4.2). Null distribution of F estimated by χ^2-approximation. Bandwidth selector for the kernel log-ratio function (6.2.3) in rows, and bandwidth selector for the kernel regression function (6.3.1) in columns.

FIGURE C.10: Power comparison for simulated patterns with intensity given in expression (6.4.2). Bandwidth selector for the kernel log-ratio in columns, and bandwidth selector for the kernel regression in rows.



Comparison with previous separability tests

FIGURE C.11: Probability of acceptance for separable patterns, and type II error for nonseparable patterns. Simulated patterns with intensity function (6.4.1). F: separability test proposed in this work, with LSCV bandwidth in $\hat{\rho}(x,t)$ and AICC bandwidth in $\hat{m}(x)$. S_3 and S_4 by Schoenberg (2004). KL and H discrepancy measures by Díaz-Avalos et al. (2013).

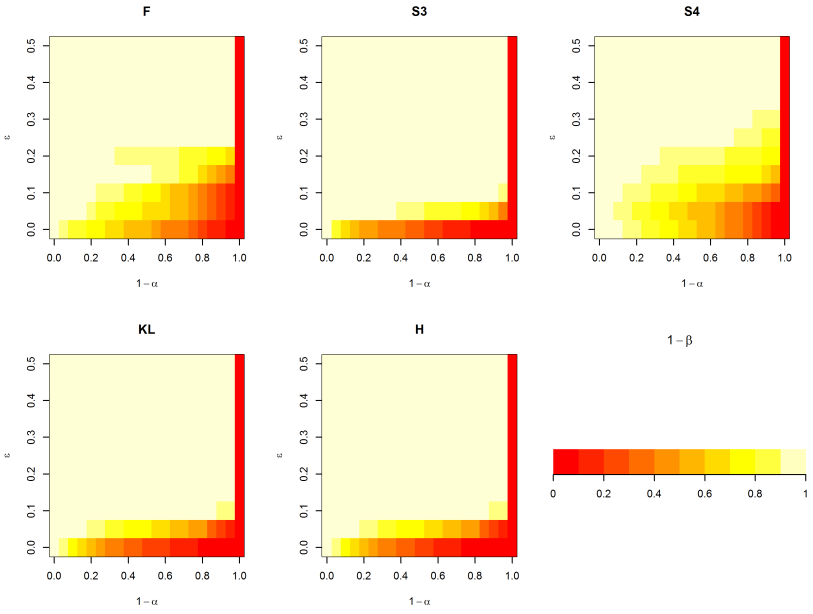
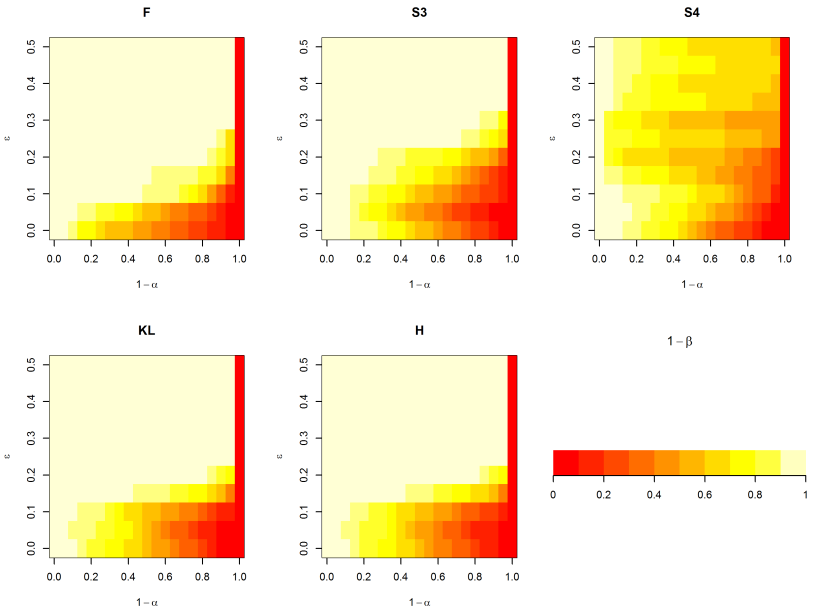


FIGURE C.12: Probability of acceptance for separable patterns, and type II error for nonseparable patterns. Simulated patterns with intensity function (6.4.1). See details in the caption of Figure C.11



Resumen en español

Introducción

Podemos encontrar conjuntos de datos distribuidos de forma irregular en una región plana en ámbitos tan importantes como son la ecología, ingeniería forestal, sismología, epidemiología, astronomía o geografía. Estos conjuntos de datos se denominan *patrones puntuales espaciales* y llamamos *evento* a cada dato para distinguirlo del resto de puntos de la región.

Un *proceso puntual* es un proceso estocástico que genera un número finito de eventos $\mathbf{X} = \{\mathbf{x_1}, \ldots, \mathbf{x_N}\}$ en una región $W \subset \mathbb{R}^d$. Si los eventos tienen asociadas medidas o marcas tenemos un *proceso puntual con marcas*. Los *procesos puntuales espaciales* generan eventos en una región plana, $W \subset \mathbb{R}^2$. Si además de la localización espacial de cada evento conocemos el momento en el que se ha producido $\mathbf{S} = \{(\mathbf{x_1}, \mathbf{t_1}), \ldots, (\mathbf{x_N}, \mathbf{t_N})\} \subset W \times T \subset \mathbb{R}^2 \times T$, tenemos una realización de un *procesos puntual espacio-temporal*. Se puede encontrar información detallada sobre el análisis de procesos puntuales en Daley y Vere-Jones (2003, 2007), Moller y Waagepetersen (2003), Illian et al. (2008) y Diggle (2013).

Las características de primer y segundo orden nos permiten definir medidas análogas a la media y la varianza en el ámbito de los procesos puntuales. La *intensidad de primer orden* de un proceso puntual es

$$\lambda(x) = \lim_{|dx| \to 0} \left\{ \frac{E\left[N(dx)\right]}{|dx|} \right\}$$

donde E denota la esperanza de una variable aleatoria, $|dx|$ y $N(dx)$ son el área y el número de eventos de $\mathbf{X}$ observados en dx. Intuitivamente, $\lambda(x)|dx|$ se puede interpretar

como la probabilidad de tener exactamente un evento de $\mathbf{X}$ en dx. Un proceso puntual es homogéneo si su intensidad es constante, $\lambda(x) = \lambda > 0$, e inhomogéneo en caso contrario.

Las características de segundo orden, que se basan en el análisis de pares de eventos, describen la estructura de dependencia del proceso. Aunque se han desarrollado diversas características de segundo orden para procesos puntuales, una de las medidas más utilizadas es la *función de segundo momento reducida* o *K-función* (Ripley, 1977), que mide el número esperado de eventos a distancia menor que r de un evento arbitrario.

La estimación de la intensidad de primer orden es uno de los puntos fundamentales en el análisis de procesos puntuales espaciales y espacio-temporales. Esta función caracteriza la distribución de los eventos en la región de observación, y además es necesaria para estimar las característica de segundo orden.. Siguiendo las ideas de estimación de densidades, Diggle (1985) propuso el estimador núcleo de la intensidad de prime orden para procesos puntuales en $\mathbb{R}$, cuya extensión a los marcos espacial y espacio-temporal es directa. Sin embargo, la falta de consistencia de este estimador ha limitado su uso al análisis exploratorio, mientras que en otras áreas de la estadística los estimadores tipo núcleo han proporcionado una gran cantidad de herramientas de inferencia y contraste. Para solucionar este problema Cucala (2006) definió la densidad de localización de eventos y demostró la consistencia de su estimador núcleo, pero este estimador no ha recibido mucha atención.

La estimación de la intensidad de un proceso puntual espacio-temporal tiene una dificultad añadida: el ajuste conjunto de las localizaciones espaciales y los tiempos de ocurrencia de los eventos. Esta dificultad aumenta cuando el proceso tiene asociada alguna marca o depende de covariables. Por este motivo, la mayoría de los modelos actuales suponen que la intensidad es separable, es decir, que se puede expresar como el producto de sus componentes espacial y temporal, lo que facilita considerablemente su estimación y evita la maldición de la dimensionalidad. Sin embargo, esta hipótesis puede ser muy restrictiva cuando analizamos datos reales. Esta situación ha motivado el desarrollo de tests de separabilidad no paramétricos que calibran la distribución del estadístico de contraste mediante simulaciones del modelo separable (Schoenberg, 2004; Assuncáo y Maia, 2007; Chang y Schoenberg, 2011; Díaz-Avalos et al., 2013).

El objetivo de esta tesis es proporcionar nuevas herramientas metodológicas para realizar inferencia no paramétrica en las características de primer orden de procesos puntuales

espaciales y espacio-temporales. Para ello se analiza bajo que condiciones podemos extender técnicas de inferencia no paramétrica desarrolladas recientemente para el análisis de datos univariantes y multivariantes al marco de los procesos puntuales. Teniendo en cuenta los problemas expuestos anteriormente, la tesis se ha centrado en los siguientes aspectos:

1. Desarrollar un procedimiento de bootstrap suavizado para procesos puntuales espaciales inhomogéneos de Poisson que nos permita estimar de forma consistente el MISE (error cuadrático medio integrado) del estimador núcleo de la densidad de localización de eventos, y proponer un criterio de selección de la matríz de suavizado del estimador núcleo de la intensidad basado en el MISE bootstrap.

2. Desarrollar un test no paramétrico para comparar las funciones de intensidad de procesos puntuales espaciales inhomogéneos.

3. Proponer un nuevo test no paramétrico de separabilidad espacio-temporal cuya calibración no dependa de la simulación de replicas del proceso separable.

4. Utilizar las técnicas de análisis de procesos puntuales disponibles actualmente y los nuevos métodos propuestos en esta tesis para estudiar el comportamiento de los incendios forestales registrados en Galicia en el periodo 1999-2008.

A continuación se exponen las principales contribuciones de la tesis a la inferencia en procesos puntuales espaciales y espacio-temporales, y se presenta el conjunto de datos reales que se ha usado para ilustrar la aplicación de los métodos desarrollados.

Contribuciones a la inferencia en procesos puntuales espaciales

Estimador consistente de la intensidad de primer orden

La estimación de la intensidad de primer orden es una de las primeras tareas a las que nos enfrentamos en el análisis de cualquier patrón puntual espacial. Esta tarea puede abordarse mediante técnicas paraméricas o no paramétricas. El ajuste paramétrico de la función de intensidad mediante máxima pseudoverosimilitud (Diggle, 2003; Waagepetersen, 2007; Illian et al., 2008) es una técnica usada habitualmente, pero puede llevarnos a estimaciones erróneas si no se especifica correctamente el modelo. La inferencia no paramétrica proporciona métodos alternativos que ajustan la intensidad a partir de los datos observados sin necesidad de asumir ningún modelo. Diggle (1985) introdujo el

estimador núcleo de la intensidad para procesos puntuales en $\mathbb{R}$, que puede extenderse de forma natural al dominio espacial. Sin embargo la falta de consistencia de este estimador ha limitado su uso al análisis exploratorio. Esta limitación puede ser también la causa de la poca atención que ha recibido en los últimos años la selección del parámetros de suavizado. De hecho resulta paradójico que, a pesar de que los procesos puntuales espaciales se observan en regiones de $\mathbb{R}^2$, se siga utilizando una ventana escalar para construir el estimador núcleo de la intensidad.

Partiendo de la propuesta de Cucala (2006), nos centramos en el estimador núcleo de la densidad de localización de eventos, $\lambda_0(x) = \lambda(x)/m$, donde $m = \int_W \lambda(x)dx$ es el número esperado de eventos del proceso. La equivalencia entre esta función y la densidad de una variable aleatoria bidimensional, nos permitirá demostrar la consistencia de su estimador núcleo:

$$\hat{\lambda}_{0,H}(x) = \frac{\hat{\lambda}_H(x)}{N} I[N \neq 0] = (p_H(x)N)^{-1} |H|^{-1/2} \sum_{i=1}^{N} k\left(H^{-1/2}\left(x - \mathbf{x_i}\right)\right) I[N \neq 0]$$

donde $\hat{\lambda}_H(x)$ es el estimador núcleo de la intensidad; la función núcleo, $k(\cdot)$, es una densidad bivariante radialmente simétrica; H es una matriz de suavizado simétrica y definida positiva, y $|H|$ denota el determinante de H; y $p_h(x) = \int_W h^{-2}k((x-y)/h)dy$ es el corrector del efecto frontera, que surge cuando observamos datos en una región acotada.

Bajo un marco asintótico de intensidad creciente "infill asymptotics", que asume que el número esperado de eventos generados por el proceso puntual tiende a infinito, y asumiendo que la intensidad de primer orden tiene derivadas parciales continuas y con cuadrado integrable hasta orden 4, se obtiene que el error cuadrático medio integrado (MISE) de $\hat{\lambda}_{0,H}(x)$ verifica $MISE(H) = AMISE(H) + o\left(A(m)|H|^{-1/2} + tr^2(H)\right)$, donde

$$AMISE(H) = \frac{1}{4}\mu_2(k)^2 (vech^T H)\Psi_4(vechH) + A(m)|H|^{-1/2}R(k)$$

$AMISE(H) \to 0$ cuando $m \to \infty$ y todos los elementos de H tienden a 0, lo que demuestra la consistencia de $\hat{\lambda}_{0,H}(x)$.

Con el fin de obtener un estimador del AMISE que nos permita utilizar este error como criterio de selección de matriz de suavizado, se extiende al dominio espacial el algoritmo de bootstrap suavizado propuesto por Cowling et al. (1996) para procesos puntuales en

$\mathbb{R}$, que genera réplicas del proceso a partir del estimador núcleo de la intensidad del proceso puntual. Bajo el mismo marco asintótico que en el paso anterior y asumiendo que $\lambda(x)$ tiene derivadas parciales continuas y con cuadrado integrable hasta orden 6, se demuestra la consistencia del AMISE bootstrap y del estimador bootstrap de la intensidad. Por último, dado que el MISE bootstrap no depende de las réplicas sino del patrón observado,vemos que no es necesario aplicar el algoritmo para calcularlo.

Los resultados anteriores, junto con el paralelismo entre la densidad de localización de eventos de un proceso puntual espacial y la función de densidad bivariante, sugieren un selector de ventana plug-in análogo al propuesto por Duong y Hazelton (2003 para datos multivariantes. Los resultados del estudio de simulación realizado para ver la validez del selector de ventana plug-in y la aplicación a datos reales avalan el uso de este criterio, que proporciona ventanas adecuadas incluso para procesos que no son de Poisson.

Comparación no paramétrica de funciones de intensidad

Cuando analizamos un procesos puntual multitipo es natural preguntarse si dos tipos de eventos tienen la misma distribución espacial. Por ejemplo, podemos preguntarnos si la distribución espacial del riesgo de incendios intencionados y con causa natural es la misma. Sin embargo, hasta el momento el análisis de procesos puntuales muttitipo se ha centrado en contrastar si existe interacción espacial entre dos tipos eventos, es decir, en las características de segundo orden.

Si dos procesos puntuales, $\mathbf{X}_1$ y $\mathbf{X}_2$, tienen la misma distribución espacial, sus intensidades de primer orden son proporcionales y sus densidades de localización de eventos son iguales. Por tanto, para comparar la distribución espacial de dos procesos puntuales necesitamos un test que compare densidades de localización de eventos. La consistencia del estimador núcleo de la densidad de localización de eventos y la equivalencia entre esta función y la función de densidad bivariante, sugieren extender al ámbito de los procesos puntuales el test desarrollado por Duong et al. (2012) para comparar distribuciones multivariantes.

Condicionados al número de eventos, $N_1 = n_1$ y $N_2 = n_2$, los patrones puntuales espaciales pueden verse como una muestras aleatoria de las distribuciones bivariantes con densidades $\lambda_{01}(\cdot)$ y $\lambda_{02}(\cdot)$. Entonces podemos contrastar la hipótesis nula, $\mathcal{H}_0$: $\lambda_{01}(x) = \lambda_{02}(x)$, mediante el siguiente estadístico

$$\hat{T} = \int_W \left(\hat{\lambda}_{01}(x) - \hat{\lambda}_{02}(x)\right)^2 dx$$
$$= \hat{\psi}_{0,1} + \hat{\psi}_{0,2} - \left(\hat{\psi}_{0,12} + \hat{\psi}_{0,21}\right)$$

donde $\psi_{0,ij} = \int_W \lambda_{0i}(x)\,\lambda_{0j}(x)\,dx$, para $j = 1, 2$, y G_1, G_2 son las matrices de suavizado de los estimadores núcleo $\hat{\psi}_{0,1}$ y $\hat{\psi}_{0,2}$.

Duong et al. (2012) prueban que $\hat{T}$ es asintoticamente normal, propiedad que se debería extender de forma natural al caso de procesos inhomogéneos de Poisson. Sin embargo sabemos que la calibración asintotica del test puede no ser adecuada para patrones espaciales con pocos eventos y, por otra parte, no podemos garantizar la normalidad de $\hat{T}$ si los procesos no son de Poisson. Por tanto, se propone calibrar el estadístico de contraste mediante un algoritmo de bootstrap suavizado que genera réplicas de la hipótesis nula. Por último, se propone utilizar algoritmos plug-in para seleccionar las matrices de suavizado de los estimadores núcleo implicados en el test.

La validez del test se ha comprobado mediante un estudio de simulación, que confirma la necesidad de calibrar el test mediante bootstrap, y que muestra su potencia para distintos tipos de desviaciones de la hipótesis nula. También se ha ilustrado la aplicación del test al análisis de incendios forestales.

Contribuciones a la inferencia en procesos puntuales espacio-temporales

Test no paramétrico de separabilidad espacio-temporal

La función de intensidad de un proceso puntual espacio-temporal

$$\lambda(x,t) = \lim_{|dx \times dt| \to 0} \left\{ \frac{E\left[N(dx,dt)\right]}{|dx \times dt|} \right\}$$

mide la tasa esperado de eventos por unidad de volumen espacio-temporal, y por tanto, caracteriza la estructura de primer orden de los procesos puntuales espacio-temporales: Por tanto su estimación es uno de los pasos fundamentales en el análisis de patrones

puntuales espacio-temporales. Dado que el ajuste conjunta de las localizaciones espaciales y los tiempos de ocurrencia de los eventos puede resultar complicado, la mayoría de los modelos disponibles asumen que el proceso puntual es separable. Esta hipótesis nos permite descomponer la intensidad en sus componentes espacial y temporal, facilitando su estimación. Sin embargo, en la aplicación a datos reales asumir que el proceso es separable puede dar lugar a modelos restrictivos y poco realistas. Esta situación pone de manifiesto la necesitad de contrastar esta hipótesis. Recientemente se han desarrollado tests no paramétricos de separabilidad que calibran la distribución del estadístico de contraste mediante simulaciones del modelo separable (Schoenberg, 2004; Assuncáo y Maia, 2007; Chang y Schoenberg, 2011; Díaz-Avalos et al., 2013).

En un procesos separable, el riesgo de observar un evento en el instante t no depende de su localización espacial, es decir el cociente entre la intensidad espacio-temporal y la intensidad de la componente espacial del proceso no depende de la localización de los eventos. Esta propiedad sugiere un nuevo test de separabilidad que contraste si el cociente $r(x,t) = \lambda(x,t)/\lambda_1(x)$ depende de las localizaciones espaciales.

Antes de implementar el nuevo test es necesario estimar $r(x,t)$. Teniendo en cuenta la similitud entre $r(x,t)$ y la función de riesgo relativo espacio-temporal (Kelsall y Diggle, 1995a, Sarojinie Fernando y Hazelton, 2014), se ha propuesto un estimador núcleo del log-ratio $\rho(x,t) = f(x,t)/g(x)$, donde f y g son las densidades de localización de eventos de $\mathbf{S}$ y $\mathbf{X}$.

$$\hat{\rho}(x,t) = log\left(\frac{\hat{f}_{h_s,h_t}(x,t)+\delta}{\hat{g}_{h_s}(x)+\delta}\right) = \log\left(\hat{f}_{h_s,h_t}(x,t)+\delta\right) - \log\left(\hat{g}_{h_s}(x)+\delta\right)$$

donde, $h_s = (h_{s1}, h_{s2}) = dgH_s^{1/2}$ es la matriz de suavizado diagonal común a la componente espacial de $f(\cdot)$ y a $g(\cdot)$, y h_t es el parámetro de suavizado de la componente temporal. Se ha comprobado la consistencia de $\rho(\hat{x},t)$ y se ha propuesto un criterio de validación cruzada para seleccionar la matriz diagonal de suavizado.

Siguiendo la propuesta de Bowman y Azzalini (1997) se aplica un contraste de regresión no paramétrico para analizar si el valor de la función log-ratio en cada evento, $Y = \{y_i = \rho(\mathbf{x_i}, \mathbf{t_i}); i = 1, \ldots, n\}$, depende de su localización espacial, $X = \{x_i = (x_{i1}, x_{i2}), i = 1, \ldots, n\}$. En los estudios de simulación realizados para analizar la validez del test se ha establecido la necesidad de aplicar un test de permutaciones para calibrar

el estadístico de contraste, se ha visto que el selector de validación cruzada de la matriz de suavizado para $\hat{\rho}(x,t)$ contribuye a la correcta calibración del test, y que el parámetro de suavizado de la función de regresión tiene un menor efecto en el comportamiento del test. Por último se ha visto que este test mejora los resultados obtenidos con los test propuestos por Schoenberg (2004) y Díaz-Avalos et al. (2013).

Datos de incendios

Los incendios forestales suponen uno de los principales problemas medioambientales a los que se enfrenta la sociedad actual. En particular, en Galicia los incendios intencionados, que durante las últimas décadas han representado más del 75% del total de incendios, son la principal causa de destrucción de los bosques y, por tanto, suponen un gran desafío tanto para el gobierno como para los gestores forestales. Esta situación ha generado un creciente interés en el desarrollo de herramientas estadísticas que nos permitan caracterizar el comportamiento de los incendios forestales.

Si asociamos a cada incendio las coordenadas del punto de ignición y el momento en que fue detectado, junto con otra información como causa del incendio o superficie quemada, estamos ante una realización un proceso puntual espacio-temporal. Las técnicas de análisis e inferencia desarrolladas recientemente para procesos puntuales espaciales y espacio-temporales nos permiten describir la distribución espacial de los incendios registrados en una determinada región, y caracterizar tanto la dependencia entre incendios como su relación con variables ambientales o socio-económicas. Durante los últimos años varios trabajos han recurrido a los procesos puntuales paa analizar y tratar de predecir el riesgo de incendios forestales (Genton et al., 2006; Schoenberg et al., 2007; Yang et al.,2007; Hering et al., 2009; Turner, 2009; Moller y Díaz-Avalos, 2010; Xu et al., 2011; Juan et al., 2012; Fuentes-Santos et al., 2013; Pereira et al., 2013; Serra et al., 2014).

La base de datos usada para ilustrar la aplicación de los procesos puntuales en esta tesis recoge los 85134 incendios forestales registrados en Galicia en el periodo 1999-2008. Estos incendios quemaron 318861 ha de suelo forestal, equivalente al 15% de la superficie forestal de Galicia, y 5014 ha de superficie rasa. Junto con las coordenadas espaciales de los puntos de ignición y el momento de inicio del fuego, se ha medido la superficie quemada por cada incendio y se han clasificados según su causa en intencionados, naturales, debidos a negligencia, reproducciones y con causa desconocida. De acuerdo con la clasificación del PLADIGA 2015 (`http://www.medioruralemar.xunta.`

es/areas/forestal/incendios_forestais/pladiga_2015), hemos definido tres tipos de incendios según su superficie quemada (S): incendios pequeños ($S < 1$ ha), medianos ($1 \leq S < 25$ ha) y grandes ($S \geq 25$ ha). Por tanto, la base de datos se puede ver como una realización de un proceso puntual espacio-temporal multitipo.

El análisis espacial de los patrones de incendios forestales nos ha permitido caracterizar su distribución espacial y las interacciones entre puntos de ignición. Los estimadores núcleo de las intensidades de primer orden muestran que el riesgo de incendios depende de la localización espacial, y nos han permitido identificar las áreas con mayor riesgo para los distintos tipos de incendios. La comparación no paramétrica de intensidades ha detectado variabilidad interanual en la distribución espacial de incendios grandes, intencionados y naturales.

El análisis de segundo orden ha detectado interacción positiva entre incendios pequeños, medianos y grandes. Este resultado concuerda con el patrón de comportamiento de los pirómanos, que intentan quemar una zona reiteradamente hasta que lo consiguen, y además sugieren que podemos usar los incendios con menos de una hectárea quemada para predecir el riesgo de incendio grande. Por otra parta, la independencia entre incendios naturales e intencionados y el hecho de que tengan distintas distribuciones espaciales sugieren que debemos usar covariables distintas para estimar el riesgo de estos dos tipos de incendios.

Estructura de la tesis

La memoria de la tesis se ha organizado como sigue:

- En el capitulo 1 se presentan los objetivos de la tesis y la base de datos de incendios forestales.
- El capitulo 2 contiene una revisión de la teoría de procesos puntuales espaciales y espacio-temporales. Se introducen sus propiedades y características de primer y segundo orden, y se proporciona información relevante para un correcto seguimiento de los capítulos siguientes.
- En el capitulo 3 se discute el estimador núcleo de la intensidad de un proceso puntual espacial y se analiza la consistencia del estimador núcleo de la densidad de localización de eventos con matriz de suavizado. A continuación se introduce un

procedimiento de bootstrap suavizado para procesos espaciales inhonogéneos, se demuestra la consistencia del MISE bootstrap, y se propone un selector de ventana plug-in. Las aportaciones de este capítulo han sido publicadas en Fuentes-Santos et al. (2015).

- En el capitulo 4 se introduce un test no paramétrico para comparar las características de primer orden de patrones puntuales espaciales y se propone un algoritmo bootstrap para estimar la distribución nula del estadístico de contraste.

- En el capítulo 5 se analiza el comportamiento de los incendios registrados en Galicia en el periodo 1999-2008 con la ayuda de las técnicas desarrolladas hasta el momento y los nuevos métodos propuestos en este trabajo. Se estudia la distribución espacial de los incendios, y la interacción espacial entre puntos de ignición. Este capítulo es una extensión a toda Galicia del análisis de incendios forestales registrados en el distrito Fonsagrada-Ancares hecho en Fuentes-Santos et al. (2013).

- El capitulo 6 trata la hipótesis de separabilidad en procesos puntuales espacio-temporales. Se propone utilizar un contraste de regresión no paramétrico para contrastar si el cociente entre la función de intensidad espacio-temporal y su componente espacial depende de la localización de los eventos, que es equivalente a contrastar si la intensidad del proceso es separable. Para implementar el test, se introduce el estimador núcleo de la función log-ratio y se propone un critero de selección de ventana.

Bibliography

Assuncáo, R. and Maia, A. (2007). A note on testing separability in spatial-temporal marked point processes. *Biometrics*, 63(1):290–294.

Baddeley, A., Møller, J., and Waagepetersen, R. (2000). Non- and semiparametric estimation of interaction in inhomogeneous point patterns. *Statistica Neerlandica*, 54(3):329–350.

Baddeley, A. and Turner, R. (2005). Spatstat: an R package for analyzing spatial point patterns. *Journal of Statistical Software*, 12(6):1–42. URL: `www.jstatsoft.org`, ISSN: 1548-7660.

Baddeley, A. J. and Silverman, B. W. (1984). A cautionary example on the use of second-order methods for analyzing point patterns. *Biometrics*, 40(4):1089.

Barr, C. D. and Schoenberg, F. P. (2010). On the voronoi estimator for the intensity of an inhomogeneous planar poisson process. *Biometrika*, 97(4):977–984.

Berman, M. and Diggle, P. (1989). Estimating weighted integrals of the second-order intensity of a spatial point process. *Journal of the Royal Statistical Society. Series B (Methodological)*, 51(1):pp. 81–92.

Besag, J. (1977). Contribution to the discussion of Dr Ripley's paper. *Journal of the Royal Statistical Society, Series B (Methodological)*, 39(2):193–195.

Bithell, J. (1990). An application of density estimation to geographical epidemiology. *Statistics in medicine*, 9(6):691–701.

Bowman, A. W. (1984). An alternative method of cross-validation for the smoothing of density estimates. *Biometrika*, 71(2):353–360.

Bowman, A. W. and Azzalini, A. (1997). *Applied Smoothing Techniques for Data Analysis: The Kernel Approach with S-Plus Illustrations*. Oxford Statistical Science Series 18.

Bowman, A. W. and Azzalini, A. (2014). *R package \textttsm: nonparametric smoothing methods (version 2.2-5.4)*. University of Glasgow, UK and Università di Padova, Italia.

Brooks, M. M. and Marron, J. S. (1991). Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications*, 38(1):157–165.

Calduch, M.-A. (2004). *Pseudolikelihood and inhomogeneity in spatial point processes*. PhD thesis, Universitat Jaume I, Castellón.

Cao, R. (1993). Bootstrapping the mean integrated squared error. *Journal of Multivariate Analysis*, 45(1):137–160.

Cao, R., Cuevas, A., and González Manteiga, W. (1994). A comparative study of several smoothing methods in density estimation. *Computational Statistics & Data Analysis*, 17(2):153–176.

Cao, R. and Lugosi, G. (2005). Goodness-of-fit Tests Based on the Kernel Density Estimator. *Scandinavian Journal of Statistics*, 32(4):599–616.

Chacón, J. E. and Duong, T. (2010). Multivariate plug-in bandwidth selection with unconstrained pilot bandwidth matrices. *Test*, 19(2):375–398.

Chacón, J. E. and Duong, T. (2011). Asymptotics for general multivariate kernel density derivative estimators. *Statistica Sinica*, 21:807–840.

Chang, C.-H. and Schoenberg, F. P. (2011). Testing separability in marked multidimensional point processes with covariates. *Annals of the Institute of Statistical Mathematics*, 63(6):1103–1122.

Comas, C., Palahi, M., Pukkala, T., and Mateu, J. (2009). Characterising forest spatial structure through inhomogeneous second order characteristics. *Stochastic Environmental Research and Risk Assessment*, 23(3):387–397.

Cowling, A., Hall, P., and Phillips, M. J. (1996). Bootstrap confidence regions for the intensity of a poisson point process. *Journal of the American Statistical Association*, 91(436):1516–1524.

Cressie, N. A. C. (1993). *Statistics for Spatial Data*. Wiley-Interscience, revised edition edition.

Cucala, L. (2006). *Espacements bidimensionnels et données entachées d´erreurs dans l´analyse des processus ponctuels spatiaux*. PhD thesis, Université des Sciences Sociales, Toulouse I.

Cucala, L. (2008). Intensity estimation for spatial point processes observed with noise. *Scandinavian Journal of Statistics*, 35(2):322–334.

Daley, D. J. and Vere-Jones, D. (2003). *An introduction to the theory of point processes: volume I elementary theory and methods*, volume 2. Springer-Verlag, New York.

Daley, D. J. and Vere-Jones, D. (2007). *An introduction to the theory of point processes: volume II: general theory and structure*, volume 2. Springer-Verlag, New York.

Davies, T. M., Hazelton, M. L., Marshall, J. C., and others (2011). Sparr: analyzing spatial relative risk using fixed and adaptive kernel density estimation in R. *Journal of Statistical Software*, 39(i01).

Devroye, L. and Györfi, L. (1985). *Nonparametric density estimation: the L1 view*. Wiley.

Díaz-Avalos, C., Juan, P., and Mateu, J. (2013). Similarity measures of conditional intensity functions to test separability in multidimensional point processes. *Stochastic Environmental Research and Risk Assessment*, 27(5):1193–1205.

Diggle, P. (1985). A kernel method for smoothing point process data. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, 34(2):pp. 138–147.

Diggle, P. (2003). *Statistical Analysis of Spatial Point Patterns*, volume 2nd ed. A Hodder Arnold Publication, Oxford, UK.

Diggle, P. and Marron, J. S. (1988). Equivalence of Smoothing Parameter Selectors in Density and Intensity Estimation. *Journal of The American Statistical Association*, 83:793–800.

Diggle, P. J. (2013). *Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition*. CRC Press.

Diggle, P. J., Chetwynd, A. G., Häggkvist, R., and Morris, S. E. (1995). Second-order analysis of space-time clustering. *Statistical methods in medical research*, 4(2):124–136.

Duong, T. (2013a). *ks: Kernel smoothing*. R package version 1.8.13.

Duong, T. (2013b). Local significant differences from nonparametric two-sample tests. *Journal of Nonparametric Statistics*, 25(3):635–645.

Duong, T., Goud, B., and Schauer, K. (2012). Closed-form density-based framework for automatic detection of cellular morphology changes. *Proceedings of the National Academy of Sciences of the United States of America*, 109(22):8382–8387.

Duong, T. and Hazelton, M. (2003). Plug-in bandwidth matrices for bivariate kernel density estimation. *Journal of Nonparametric Statistics*, 15(1):17–30.

Duong, T. and Hazelton, M. L. (2005). Cross-validation bandwidth matrices for multivariate kernel density estimation. *Scandinavian Journal of Statistics*, 32(3):485–506.

Fan, Y. (1994). Testing the goodness of fit of a parametric density function by kernel method. *Econometric Theory*, 10:316–316.

Flannigan, P. and Wotton, B. (2001). *Forest Fires: Behavior and Ecological Effects*, chapter Climate, weather and area burned, pages 351–373. Academic Press.

Fuentes-Santos, I., González-Manteiga, W., and Mateu, J. (2015). Consistent smooth bootstrap kernel intensity estimation for inhomogeneous spatial Poisson point processes. *Scandinavian Journal of Statistics*. DOI: 10.1111/sjos.12183.

Fuentes-Santos, I., Marey-Pérez, M., and González-Manteiga, W. (2013). Forest fire spatial pattern analysis in Galicia (NW Spain). *Journal of environmental management*, 128:30–42.

Gabriel, E. and Diggle, P. J. (2009). Second-order analysis of inhomogeneous spatio-temporal point process data. *Statistica Neerlandica*, 63(1):43–51.

Gabriel, E., Rowlingson, B. S., and Diggle, P. J. (2013). stpp: An R Package for Plotting, Simulating and Analyzing Spatio-Temporal Point Patterns. *Journal of Statistical Software*, 53(2):1–29.

Genton, M. G., Butry, D. T., Gumpertz, M. L., and Prestemon, J. P. (2006). Spatio-temporal analysis of wildfire ignitions in the St Johns River water management district, Florida. *International Journal of Wildland Fire*, 15(1):87–97.

González-Manteiga, W., Cao, R., and Marron, J. S. (1996). Bootstrap selection of the smoothing parameter in nonparametric hazard rate estimation. *Journal of the American Statistical Association*, 91(435):1130–1140.

Guan, Y. (2007). A composite likelihood cross-validation approach in selecting bandwidth for the estimation of the pair correlation function. *Scandinavian Journal of Statistics*, 34(2):336–346.

Guan, Y. (2008a). A goodness-of-fit test for inhomogeneous spatial poisson processes. *Biometrika*, 95(4):831–845.

Guan, Y. (2008b). On consistent nonparametric intensity estimation for inhomogeneous spatial point processes. *Journal of the American Statistical Association*, 103(483):1238–1247.

Guan, Y. and Loh, J. M. (2007). A thinned block bootstrap variance estimation procedure for inhomogeneous spatial point patterns. *Journal of the American Statistical Association*, 102(480):pp. 1377–1386.

Guttorp, P. and Minin, V. N. (1995). *Stochastic Modeling of Scientific Data*. CRC Press.

Hazelton, M. L. and Davies, T. M. (2009). Inference based on kernel estimates of the relative risk function in geographical epidemiology. *Biometrical Journal*, 51(1):98–109.

Hazelton, M. L. and Marshall, J. C. (2009). Linear boundary kernels for bivariate density estimation. *Statistics & Probability Letters*, 79(8):999–1003.

Hering, A. S., Bell, C. L., and Genton, M. G. (2009). Modeling spatio-temporal wildfire ignition point patterns. *Environmental and Ecological Statistics*, 16(2):225–250.

Hope, A. (1968). A simplified monte carlo significance test procedure. *Journal of the Royal Statistical Society. Series B*, 30:582–598.

Hurvich, C. M., Simonoff, J. S., and Tsai, C. L. (1998). Smoothing parameter selection in nonparametric regression using an improved Akaike information criterion. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(2):271–293.

Illian, D. J., Penttinen, P. A., Stoyan, D. H., and Stoyan, D. (2008). *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons.

Johnson, E. A. (1996). *Fire and Vegetation Dynamics: Studies from the North American Boreal Forest*. Cambridge University Press.

Juan, P., Mateu, J., and Saez, M. (2012). Pinpointing spatio-temporal interactions in wildfire patterns. *Stochastic Environmental Research and Risk Assessment*, 26(8):1131–1150.

Kagan, Y. (1999). Universality of the seismic moment-frequency relation. *Pure Applyed Geophysics*, 155:233–258.

Kelsall, J. E. and Diggle, P. J. (1995a). Kernel estimation of relative risk. *Bernoulli*, 1:3–16.

Kelsall, J. E. and Diggle, P. J. (1995b). Non-parametric estimation of spatial variation in relative risk. *Statistics in Medicine*, 14:2335–2342.

Lahiri, S. N., Kaiser, M. S., Cressie, N., and Hsu, N.-J. (1999). Prediction of spatial cumulative distribution functions using subsampling. *Journal of the American Statistical Association*, 94(445):86–97.

Loader, C. (1999). *Local Regression and Likelihood*. Statistics and Computing. Springer, New York.

Loh, J. and Stein, M. (2004). Bootstrapping a spatial point process. *Statistica Sinica*, 14(1):69–102.

Loh, J. M. (2010). Bootstrapping an inhomogeneous point process. *Journal of Statistical Planning and Inference*, 140(3):734 – 749.

Loh, J. M. and Jang, W. (2010). Estimating a cosmological mass bias parameter with bootstrap bandwidth selection. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 59(5):761–779.

Marriott, F. H. C. (1979). Barnard's monte carlo tests: How many simulations? *Applied Statistics*, 28(1):75.

Marron, J. S. and Wand, M. P. (1992). Exact mean integrated squared error. *The Annals of Statistics*, 20(2):712–736.

Marshall, J. C. and Hazelton, M. L. (2010). Boundary kernels for adaptive density estimators on regions with irregular boundaries. *Journal of Multivariate Analysis*, 101(4):949–963.

Møller, J. and Díaz-Avalos, C. (2010). Structured Spatio-Temporal Shot-Noise Cox Point Process Models, with a View to Modelling Forest Fires. *Scandinavian Journal of Statistics*, 37(1):2–25.

Moller, J. and Waagepetersen, R. P. (2003). *Statistical inference and simulation for spatial point processes*. Chapman and Hall/CRC, Florida.

Nadaraya, E. A. (1964). On estimating regression. *Theory of Probability & Its Applications*, 9(1):141–142.

National Interagency Fire Center. (2011). .*The science of wildland fire*.

Ogata, Y. (1988). Statistical Models for Earthquake Occurrences and Residual Analysis for Point Processes. *Journal of the American Statistical Association*, 83(401):9–27.

Ogata, Y. and Katsura, K. (1993). Analysis of temporal and spatial heterogeneity of magnitude frequency distribution inferred from earthquake catalogues. *Geophysical Journal International*, 113(3):727–738.

Pereira, P., Turkman, K. F., Turkman, M. A. A., Sá, A., and Pereira, J. M. (2013). Quantification of annual wildfire risk; A spatio-temporal point process approach. *Statistica*, 73(1):55–68.

Pyne, S. J., Andrews, P. L., Laven, R. D., et al. (1996). *Introduction to wildland fire*. Number Ed. 2. John Wiley and Sons.

R Core Team (2014). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Ripley, B. (1977). Modelling spatial patterns (with discussion). *Journal of the Royal Statistical Society, series B*, 39:172–212.

Ripley, B. (1981). *Spatial Statistics*. Wiley, New York.

Rudemo, M. (1982). Empirical choice of histograms and kernel density estimators. *Scandinavian Journal of Statistics*, pages 65–78.

Sarojinie Fernando, W. T. P. and Hazelton, M. L. (2014). Generalizing the spatial relative risk function. *Spatial and Spatio-temporal Epidemiology*, 8:1–10.

Schoenberg, F., Chang, H.-C., Keeley, J., Pompa, J., Woods, J., and Xu, H. (2007). A critical assessment of the Burning Index in Los Angeles County, California. *International Journal of Wildland Fire*, 16(4):473–483.

Schoenberg, F. P. (2003). Multidimensional Residual Analysis of Point Process Models for Earthquake Occurrences. *Journal of the American Statistical Association*, 98(464):789–795.

Schoenberg, F. P. (2004). Testing separability in spatial-temporal marked point processes. *Biometrics*, 60(2):471–481.

Scott, D. (1992). *Multivariate Density Estimation. Theory, Practice and Visualization. New York: Wiley.* New York: Wiley.

Serra, L., Saez, M., Juan, P., Varga, D., and Mateu, J. (2014). A spatio-temporal Poisson hurdle point process to model wildfires. *Stochastic Environmental Research and Risk Assessment*, 28(7):1671–1684.

Shitan, M. and Brockwell, P. J. (1995). An asymptotic test for separability of a spatial autoregressive model. *Communications in Statistics - Theory and Methods*, 24(8):2027–2040.

Silverman, B. (1986). *Density Estimation for Statistics and Data Analysis (Monographs on Statistics and Applied Probability)*. Chapman and Hall/CRC.

Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika*, 76(4):705.

Turner, R. (2009). Point patterns of forest fire locations. *Environmental and ecological statistics*, 16(2):197–223.

Ushakov, N. and Ushakov, V. (2009). Some Inequalities for the Mean Integrated Squared Error of Multivariate Kernel Density Estimators. *Proceedings of the International Conference on Mathematical Methods in Reliability*, pages 376–379.

Waagepetersen, R. P. (2007). An estimating function approach to inference for inhomogeneous neyman-scott processes. *Biometrics*, 63(1):252–258.

Wand, M. and Jones, M. (1994). Multivariate plug–in bandwidth selection. *Computational Statistics 9: 97–911 (1994)*, 9:97–116.

Wand, M. P. (1992). Error analysis for general multtvariate kernel estimators. *Journal of Nonparametric Statistics*, 2(1):1–15.

Wand, M. P. and Jones, M. C. (1995). *Kernel Smoothing*. CRC Press.

Watson, G. S. (1964). Smooth regression analysis. *Sankhy: The Indian Journal of Statistics, Series A*, pages 359–372.

Xu, H., Schoenberg, F. P., and others (2011). Point process modeling of wildfire hazard in Los Angeles County, California. *The Annals of Applied Statistics*, 5(2A):684–704.

Yang, J., He, H. S., Shifley, S. R., and Gustafson, E. J. (2007). Spatial patterns of modern period human-caused fire occurrence in the Missouri Ozark Highlands. *Forest Science*, 53(1):1–15.